B85

PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [5] : C12N 15/63, 15/09, C12P 21/00	**A1**	(11) International Publication Number: **WO 94/25609** (43) International Publication Date: 10 November 1994 (10.11.94)

(54) Title: METHOD FOR CREATING OPTIMIZED REGULATORY REGIONS AFFECTING PROTEIN EXPRESSION AND PROTEIN TRAFFICKING

(57) Abstract

This invention discloses a method for optimizing the production of a polypeptide in a cell comprising the steps of identifying a regulatory region within a nucleic acid sequence to be mutagenized; preparing a nucleic acid vector comprising that region and encoding at least one polypeptide regulated by said regulatory region; deleting that region from the vector; producing a pool of random oligonucleotides; using PCR to introduce one random oligonucleotide into the position previously occupied by the regulatory region into each of a plurality of vectors to generate a pool of mutagenized vectors; introducing the mutagenized vectors into a cell sample; assaying for the expression of the polypeptide in that sample; and selecting and isolating those cells exhibiting optimized polypeptide expression. The regulatory regions contemplated within the scope of this invention include signal and protein trafficking sequences, ribosome binding sites, promoters, and translational regulatory sequences.

-1-

METHOD FOR CREATING OPTIMIZED REGULATORY REGIONS AFFECTING PROTEIN EXPRESSION AND PROTEIN TRAFFICKING

This invention relates generally to methods for
5   the mutagenesis of nucleic acid sequences and more
specifically to mutagenesis strategies involving polymerase
chain reaction-related technologies and to optimized
regulatory sequences generated by these methods.

Gene expression can be regulated at various steps
10 on the path from DNA to RNA to mature protein. The overall
path is the same for prokaryotes and eukaryotes with
regulation possible at the level of transcription of mRNA
from the DNA template, translation of mRNA at the ribosome,
and targeting of protein to appropriate locations within
15 the cell or outside. However, the differences between
prokaryotes and eukaryotes lead to a number of differences
in how gene expression can be regulated in each.
Intracellular membranes and organelles are critical
components of eukaryotic cells, but are not found in
20 prokaryotes. On the other hand, the potential for coupled
transcription and translation in prokaryotes is not
possible in compartmentalized eukaryotic cells. Eukaryotic
genes have introns interrupting the coding sequences,
prokaryotic genes do not. On the other hand, prokaryotes
25 put coding sequences for multiple genes under a single
promoter and transcribe them as a single polycistronic
mRNA, but eukaryotes do not. Most eukaryotic genes require
RNA splicing, capping, poly-adenylation, and transport from
the nucleus to the cytoplasm before the mRNA becomes
30 functional, but prokaryotes do not. In addition,
eukaryotes require a more complicated process for protein
targeting to distinguish proteins destined for different
organelles, the cell membrane, and for secretion, as well
as post translational modification of many eukaryotic
35 proteins before they can be assembled or become functional.

SUBSTITUTE SHEET (RULE 26)

-2-

In **EUKARYOTES**, the expression of a particular
gene can be regulated at transcription, RNA processing, RNA
transport, translation, protein targeting, and protein
activation/modification. Specific gene sequences are
5   responsible for effects at each level and these can be
modified to increase expression according to this
invention, if eukaryotic expression is desired.

At the level of transcription, these sequences
act directly, by interacting with DNA binding proteins.
10   Examples include: **Promoters** which interact with
transcription factors and **RNA** polymerase to initiate
transcription. **Upstream promoter elements** which are
generally near to the promoter and interact with DNA
binding proteins to either improve or inhibit the ability
15   of polymerase to interact with the promoter. The
transcription of the genes encoding the DNA binding
proteins described above and below is another level of
regulation of specific eukaryotic genes. **Enhancers**, which
are quite distant from the promoter, bind to other DNA
20   binding proteins to enhance the level of transcription of
the gene, although specific sequences are also present in
enhancers which can inhibit transcription under certain
conditions or in specific cell types. One skilled in the
art can readily determine whether transcription is
25   increased or decreased by specific sequences within
upstream promoter elements or enhancers in a specific cell
type. Thus, enhancers, promoters and upstream promoter
elements can be modified to increase expression according
to this invention.

30   At the level of **RNA processing**, primary
transcripts are capped at their 5' ends, poly-adenylated at
their 3' ends and have their introns spliced out (sometimes
leading to different protein products due to **alternative
RNA splicing**) prior to being transported through pores in
35   the nuclear membrane to cytoplasm, where they associate

SUBSTITUTE SHEET (RULE 26)

-3-

with ribosomes. Important RNA sequences which can affect
the efficiency of this process include the poly-A addition
site, the Cap site, and the splice acceptor and donor
sites, and DNA sequences encoding these sites can be
5   modified to increase expression according to this
invention.

     To participate in translation, a eukaryotic mRNA
must be transported out of the nucleus. Specific
recognition signals on the mRNA are believed necessary for
10   transit of mRNA through the nuclear pore. Other signals
may retain certain mRNAs within the nucleus, where they are
degraded. As many as half of the unspliced heterologous
nuclear RNAs produced from primary transcripts are degraded
without ever leaving the nucleus, offering a potential for
15   regulation at another level, the level of **RNA transport**.
DNA sequences encoding these RNA signals could be modified
to increase expression according to this invention as well.

     Not all mRNAs that escape the nucleus are
translated into protein. Eukaryotic mRNAs vary in
20   susceptibility to degradation. Binding to ribosomes
decreases their degradation. Important sequences that can
be modified to control expression at the level of
translation are those recognized by specific **translation
repressor proteins**, which bind to the 5' end of some mRNAs
25   and block translation initiation. For those mRNAs which
are not blocked, current data suggests that initiation of
translation is by the so-called "scanning model" of
initiation. A ribosome component (40S) binds initially to
the 5' end of an mRNA then migrates 3' until it finds an
30   initiation site where the other portion of the ribosome
joins it to form the 80S initiation complex. The Kozak
sequences surrounding the translational start site
determine whether a particular mRNA will be efficiently
translated. The presence of an Adenine (A) at the -3
35   position has as much as a twenty-fold impact on

-4-

translation. In general, the translation rate is most
affected by the nucleotide three prior to the initiator
codon and the one immediately after it [best if a Guanine
(G)]. The DNA sequence encoding binding sites for
5   translation repressor proteins and for the region around
the initiator can be modified to increase expression
according to this invention.

At the level of **protein targeting**, eukaryotic
proteins have a diversity of target locations within the
10   cell including the nucleus, nuclear membrane, cytoplasm,
mitochondria, mitochondrial inner membrane, mitochondrial
outer membrane, endoplasmic reticulum, golgi apparatus,
lysosomes, lysosomal membranes, endoplasmic reticulum
membrane, and cell membrane. In addition, secreted
15   proteins are targeted to the exterior of the cell. The
process begins, as for prokaryotes, with a signal sequence
directing the nascent peptide and its ribosomes to a
membrane. But in this case, the membrane is the rough
endoplasmic reticulum (RER). If the protein is to be
20   secreted, it is translated into the RER and begins a
journey through different membrane bound organelles and
vesicles to the cell membrane where it is extruded to the
outside by exocytosis. Proteins destined to be membrane-
bound are also translated into the RER, but do not
25   completely transit the rough endoplasmic reticulum membrane
(RERM), becoming integrated into it. This membrane fuses
with that of the Golgi apparatus or of vesicles which, in
turn, bud off and fuse with the destination membrane where
the protein either becomes bound or transits further to its
30   final target. Targeting to the proper site is almost
certainly facilitated by identifiable target signals within
the sequence of the protein that direct the protein to the
proper destination. Thus, the sequences that can be
modified to increase expression according to this invention
35   include those encoding the signal sequence, transmembrane

SUBSTITUTE SHEET (RULE 26)

-5-

sequences, cytoplasmic anchor sequences, and the putative
specific membrane targeting sequences.

In addition, numerous secreted proteins,
including immunoglobulins, consist of multiple chains that

5   must be assembled prior to secretion. In immunoglobulins,
for example, the heavy chains cannot be secreted if they
are not bound to light chains. The heavy chains that do
not bind to light chains, either because of improper
folding or an absence of free light chains, remain bound to

10 a protein in the RER called BiP, which prevents their
transit to the Golgi apparatus. Thus, it is clear that
additional regulatory elements that can be modified
according to the method of this invention to increase
expression are those which encode structural features that

15 facilitate or impede correct folding of a peptide or its
association with other chains necessary to form the mature
protein product.

At the level of **protein activation/modification**,
many eukaryotic proteins are synthesized as inactive or

20 partially active precursors that become activated by
proteolytic cleavage or other modifications (e.g.,
phosphorylation). Although, not required for expression in
many cases, secretion cannot occur without proteolytic
cleavage of the signal peptide. Inefficiency in this

25 process can delay or block secretion of functional protein.
DNA sequences encoding proteolytic cleavage sites,
phosphorylation sites, glycosylation sites, or other sites
for post-translational modification can be modified
according to this invention to yield increased expression

30 of functional protein.

In **PROKARYOTES**, there are fewer stages where
regulation of gene expression can occur, and prokaryote
specific elements and associated proteins are involved.
Regulation does take place at the levels of transcription,

35 translation, and protein targeting, as for eukaryotic

SUBSTITUTE SHEET (RULE 26)

-6-

genes, but RNA processing and RNA transport are totally
absent in prokaryotes. There are also far fewer known
cases of post-translational modification of proteins in
bacteria than in eukaryotes. Of special importance, there

5   are very few authenticated cases of glycosylated bacterial
proteins, although proteins are modified in other ways
(e.g., fatty acylation). No clear picture of regulation at
the level of post-translational modification has emerged.
Like eukaryotes, prokaryotes offer a number of specific DNA

10 sequences which can be modified according to this invention
to increase expression.

      At the level of transcription in prokaryotes
identifiable DNA sequences interact with DNA binding
proteins. The **promoter** binds RNA polymerase and, with the

15 release of the σ subunit, initiation of transcription
occurs. Unlike in eukaryotes, distant enhancer sequences
have not been seen in prokaryotes. Different promoter
sequences have different affinities for RNA polymerase.
Transcription initiation can be positively regulated by

20 binding of a catabolite gene activator protein (CAP) to a
site immediately upstream of the promoter and, in many
genes, can be negatively regulated by binding of a
repressor protein to a site immediately downstream of the
promoter called the operator. The level of transcription

25 of these DNA binding proteins can affect the
transcriptional level of the regulated gene. As one
skilled in the art will appreciate, binding sites can be
modified both in the target gene and in the gene encoding
the repressor or CAP using the present invention to

30 increase or decrease the level of transcription of
prokaryotic genes.

      Sites amenable to modification to increase
expression at the level of **translation** also exist in
prokaryotic DNA, but are distinct from those in eukaryotes.

35 A region originally known as the Shine and Dalgarno

-7-

sequence or more recently as the ribosome binding site
found on the 5' end of prokaryotic mRNAs is complementary
to the 16S ribosomal RNA. Similar to eukaryotic
initiators, an Adenine (A) is preferred at the -3 position.
5   The Shine and Dalgarno interaction is unique to prokaryotes
and is unlike the scanning model used in eukaryotes to
initiate translation described above. The DNA sequence
which encodes the ribosome binding site in mRNA can be
modified according to this invention to increase
10   expression.
         There are proteins which bind to sites on
specific mRNAs to inhibit their translation (e.g.,
ribosomal proteins). In addition, transcription can be
affected by 5' coding sequences which can lead to premature
15   termination of transcription based on charged tRNA
availability. This system, called **attenuation**, is seen in
genes for amino acid biosynthetic pathways, but is unique
to prokaryotes because it requires coupled transcription
and translation, which can occur only in cells having no
20   nuclear membrane separating the two processes. Placement
of the gene within the operon also has an effect on
translation of specific genes. The 5' genes are more
efficient in initiation than those at the 3' end of the
polycistronic message. The "polarity effect" is amplified
25   by the fact that binding to ribosomes actually delays mRNA
degradation. The DNA sequences which encode protein
sequences responsible for attenuation can be modified
according to this invention to increase expression.
         At the level of **protein targeting**, as with
30   eukaryotes, specific targeting sequences are required for
different types of targeting. These sequences encode
portions of the protein with hydrophobic or hydrophilic
properties corresponding to the protein's point of
association with the membrane. The most common sequence is
35   the N-terminal signal sequence, which is responsible for

SUBSTITUTE SHEET (RULE 26)

-8-

targeting the protein and its associated ribosomes to the
cell membrane. Internal sequences determine if portions of
the protein will reside within or outside of the lipid
bilayer of the membrane or if the protein will be secreted
5   into the periplasmic space (in gram negative) or directly
into the media (in gram positive bacteria).

        The efficiency of secretion depends upon the
signal sequence as well as the cellular machinery required
for transport. This includes peptidases for removing the
10   signal peptide and bacterial cell membrane proteins, both
of which have been implicated as potential signal peptide
receptors. A Signal Recognition Particle (SRP) analogue
has recently been confirmed for bacteria (Luirink J., et
al. Nature, 359:741-743, 1992. Another significant
15   difference between prokaryotes and eukaryotes is the
limited number of target membranes in bacteria.

        Due to all of these differences in the mechanisms
of expression between eukaryotes and prokaryotes
heterologous expression of eukaryotic genes in prokaryotic
20   cells faces a number of hurdles. First, all introns must
be removed or truncated products will be produced. Second,
eukaryotic **upstream promoter elements** and **enhancers** are not
only non-functional in prokaryotes but can be deleterious.
Third, eukaryotic signals, while occasionally functional,
25   are never optimized for prokaryotes, and often are totally
ineffective. Often fortuitous prokaryote-like signal
sequences are used that are distinct from the native
eukaryotic signal. Fourth, the absence of functional
prokaryotic ribosome binding sequences often leads to
30   little or no translation of the eukaryotic message. Fifth,
the absence of prokaryotic codon usage can slow
translation. Finally, the DNA sequence encoding an mRNA
for a eukaryotic signal could be completely misinterpreted
in a prokaryotic expression system. For example, it could
35   be recognized by a prokaryotic DNA binding protein, or the

SUBSTITUTE SHEET (RULE 26)

-9-

mRNA could experience inhibitory folding or be more rapidly degraded than in the native cell. Because such miscommunication depends in part upon the sequence of the gene to be expressed, it is possible that different genes

5   inserted into the same vector in the same heterologous expression cell will attain different levels of expression. For example, a leader sequence optimized to attain the highest level of expression of one eukaryotic protein gene in *E. coli* may not be the leader sequence that will

10   optimize expression of a different eukaryotic protein gene in *E. coli*.

The present invention overcomes these difficulties by providing a means to customize heterologous expression for any particular desired protein. The present

15   method assays the combined effect of mutations made in the DNA of one or more particular control sequences upon all steps taken during the whole process of heterologous expression leading up to secretion of a particular desired protein. Therefore, the present method provides a means to

20   customize modifications to discrete sequences in eukaryotic DNA which impact expression of the desired protein in the heterologous host cell.

For a general review of the differences in eukaryotic and prokaryotic gene expression and protein

25   transport see: Watson, JD, et al. <u>Molecular Biology of the Gene</u> 4th edition. Menlo Park, Benjamin Cummings Press, pp. 359-618 (1987); Alberts, et al. <u>Molecular Biology of the Cell</u> 2nd ed. New York, Garland Publ., pp. 551-612 (1989); Reznikoff, W. and Gold, L. eds. <u>Maximizing Gene Expression</u>

30   Boston, Butterworths Publ., pp. 1-34, 195-286 (1986); Pugsley, A.P. <u>Protein Targeting</u> San Diego, Academic Press, pp. 1-168 (1989) and Watson, et al. <u>Recombinant DNA: A Short Course</u> New York, Freeman Publ. pp. 45-57. For a review of protein translocation across the *E. coli* plasma

-10-

membrane see Wichner, et al. *Annu. Rev. Biochem.* 60:101-24,
1991 and Bieker, et al. *Trends in Gen.* 6:329-333, 1990.

The references listed in this application are
incorporated herein by reference to the extent that they
5   supplement, explain, or provide a background for each
methodology, technique and/or composition employed herein.

This invention relates to methods for optimizing
the expression of polypeptide in a cell and to methods for
creating and isolating novel regulatory sequences for the
10   expression of polypeptide in a cell. The regulatory
sequences contemplated within the scope of this invention
include, but are not limited to, signal sequences, ribosome
binding sites, promoter sequences, translational regulatory
sequences, transcription regulatory sequences, protein
15   trafficking sequences, enhancer sequences, and the like.

In a preferred aspect of the invention, a method
is provided for optimizing the production of polypeptide in
a cell comprising the steps of (a) identifying at least one
regulatory region within a nucleic acid sequence to be
20   mutagenized; (b) preparing a nucleic acid vector comprising
the regulatory region and a nucleic acid sequence encoding
at least one polypeptide regulated by the regulatory
region; (c) deleting the regulatory region from the vector;
(d) producing a pool of random oligonucleotides; (e) using
25   a polymerase chain reaction to introduce at least one
random oligonucleotide into the position previously
occupied by the regulatory region in a plurality of vectors
to generate a pool of mutagenized vectors; (f) introducing
the mutagenized vectors into a cell sample; (g) assaying
30   for the expression of the polypeptide in the cell sample;
(h) selecting cells exhibiting optimized polypeptide
expression; and (i) isolating optimized polypeptide from
the cells of step (h). Preferably the regulatory region is
located within the nucleic acid sequence encoding
35   polypeptide and in one preferred embodiment the regulatory

SUBSTITUTE SHEET (RULE 26)

-11-

region is in a translated portion of the nucleic acid
sequence encoding polypeptide. In another preferred
embodiment, the regulatory region is outside of the nucleic
acid sequence encoding polypeptide. The method is useful

5   for optimizing regulatory regions selected from the group
consisting of signal sequences, ribosome binding sites,
promoter sequences, translational regulatory sequences,
transcription regulatory sequences and protein trafficking
sequences. In a preferred embodiment of this invention,

10   the nucleic acid sequence encodes an antibiotic resistant
gene and the selection step consists essentially of growing
the cell sample in the presence of an antibiotic. In
another aspect of this embodiment, the nucleic acid
sequence encoding polypeptide encodes a selectable marker

15   and in yet another aspect of this embodiment, the nucleic
acid sequence encoding polypeptide encodes a fusion
protein. In a preferred embodiment the nucleic acid
sequence encoding polypeptide is derived from a eukaryotic
cell and the cell sample is prokaryotic.

20   In a preferred method for optimizing the
production of polypeptide in a cell, the method
additionally comprises the steps of introducing the nucleic
acid vector into a second cell sample, expressing the
polypeptide encoded by the vector in the second cell

25   sample, and measuring the level of polypeptide expression
in the second cell sample. This method optionally includes
the step of selecting cells from the first cell sample that
exhibit optimized polypeptide expression relative to the
measuring step. Once optimized cells are selected, it is

30   contemplated that the nucleic acid sequence of the random
oligonucleotide corresponding to the mutagenized regulatory
region producing optimized polypeptide expression can be
identified.

         In another preferred embodiment of this

35   invention, a method is provided for creating and isolating

SUBSTITUTE SHEET (RULE 26)

-12-

novel signal sequences comprising the steps of: (a) identifying a signal sequence within a nucleic acid sequence encoding a polypeptide; (b) preparing a nucleic acid vector comprising the nucleic acid sequence encoding

5 the polypeptide; (c) introducing the vector into a first cell sample and expressing the polypeptide in the first cell sample; (d) measuring the level of polypeptide expression in the first cell sample; (e) deleting the signal sequence from the vector; (f) producing a pool of

10 random oligonucleotides; (g) using a polymerase chain reaction to introduce at least one of the random oligonucleotides into the position previously occupied by the signal sequence in a plurality of vectors to generate a pool of mutagenized vectors; (h) introducing the

15 mutagenized vectors into a second cell sample; (i) assaying for the expression of the polypeptide in the second cell sample; (j) selecting cells exhibiting optimized polypeptide expression relative to step (d); and (k) determining the nucleic acid sequence of the regulatory

20 region contained in the mutagenized vector located in the cells of step (j). In one embodiment the optimized level of polypeptide expression is a level of expression greater than or equal to the level of expression of the polypeptide in step (d).

25   Preferably the nucleic acid vector encodes an antibiotic resistant gene and the selection step additionally consists essentially of growing the cell sample in the presence of an antibiotic. The nucleic acid sequence encoding polypeptide preferably encodes a

30 selectable marker and the polymerase chain reaction is preferably an enzymatic inverse polymerase chain reaction. In one embodiment the nucleic acid sequence encoding polypeptide encodes a fusion polypeptide and in a particularly preferred embodiment a portion of the fusion

35 protein is derived from β-lactamase. In another embodiment

SUBSTITUTE SHEET (RULE 26)

-13-

a portion of the fusion protein is derived from an
antibody. Preferably the nucleic acid sequence encoding
polypeptide is derived from a eukaryotic cell and the first
and second cell samples are prokaryotic. Still more
5   preferably, the cell samples are *E. coli*.

In yet another preferred embodiment of the method
for creating and isolating novel signal sequences, the
random oligonucleotides are biased and preferably contain
at least one positively charged amino acid at the N-
10   terminus, a stretch of at least 8 hydrophobic amino acids
and a small amino acid such as alanine, glycine, or valine
positioned at the C-terminus.

In another preferred method of this invention, a
method is disclosed for optimizing polypeptide expression
15   in a cell by performing random mutagenesis on a regulatory
region regulating polypeptide expression from a nucleic
acid vector wherein the improvement comprises deleting a
region of nucleic acid to be mutagenized, isolating the
vector containing the deletion and replacing the region to
20   be mutagenized with a random nucleic acid sequence.

In a preferred method of this invention a method
is provided for creating and isolating novel ribosome
binding sites, comprising the steps of: (a) identifying a
ribosome binding site within a nucleic acid sequence; (b)
25   preparing a nucleic acid vector comprising the nucleic acid
sequence containing a ribosome binding site operably linked
to a nucleic acid sequence encoding a polypeptide; (c)
introducing the vector into a first cell sample and
expressing the polypeptide in the first cell sample; (d)
30   measuring the level of polypeptide expression in the first
cell sample; (e) deleting the nucleic acid sequence
containing the ribosome binding site from the vector; (f)
producing a pool of random oligonucleotides; (g) using a
polymerase chain reaction to introduce at least one of the
35   random oligonucleotides into the position previously

SUBSTITUTE SHEET (RULE 26)

-14-

occupied by the nucleic acid sequence containing the
ribosome binding site in a plurality of vectors to generate
a pool of mutagenized vectors; (h) introducing the
mutagenized vectors into a second cell sample; (i) assaying
5   for the expression of the polypeptide in the second cell
sample; (j) selecting cells exhibiting optimized
polypeptide expression relative to step (d); and
determining the nucleic acid sequence containing the
ribosome binding site located in the mutagenized vector
10 introduced into the cells of step (j). In a preferred
embodiment of this method the nucleic acid sequence encodes
an antibiotic resistance gene and the selection step
consists essentially of growing the cell samples in the
presence of an antibiotic. Preferably, the nucleic acid
15 sequence encoding polypeptide encodes a selectable marker.
In one embodiment the nucleic acid sequence encoding
polypeptide is a fusion protein and preferably the fusion
protein is derived from an antibody and in another
preferred embodiment the fusion protein is derived from
20 β-lactamase.

In another preferred aspect of this invention,
polypeptide signal sequences are disclosed. These signal
sequences correspond to SEQ ID NO: 13, SEQ ID NO: 16 and
SEQ ID NO: 14 and to polypeptide signal sequences
25 containing at least contiguous amino acid 10-mers thereof.
In another preferred aspect of this invention, nucleic acid
sequences are disclosed corresponding to ribosome binding
sites identified as SEQ ID NO: 35, SEQ ID NO: 37, and SEQ
ID NO: 38 and to ribosome binding sites corresponding to at
30 least contiguous 5-mers thereof.

In yet another preferred embodiment of this
invention, a method is provided for creating and isolating
novel signal sequences, comprising the steps of: (a)
identifying a signal sequence within a nucleic acid
35 sequence encoding a polypeptide; (b) preparing a nucleic

SUBSTITUTE SHEET (RULE 26)

-15-

acid vector comprising the nucleic acid sequence encoding a
polypeptides, wherein the polypeptide is a fusion protein
having a C-terminus corresponding to β-lactamase; (c)
deleting the signal sequence from the vector; (d) producing
5   a pool of random oligonucleotide suitable for an enzymatic
inverse polymerase chain reaction; (e) using an enzymatic
inverse polymerase chain reaction to introduce the random
oligonucleotides into the position in the vector previously
occupied by the signal sequence to generate a pool of
10   mutagenized vectors; (f) introducing the mutagenized
vectors into a first cell sample of E. coli; (g) assaying
for the expression of the fusion protein in a first cell
sample in the presence of ampicillin; and (h) selecting
cells exhibiting optimized polypeptide expression. In one
15   embodiment of this method, the method additionally
comprises the steps of introducing the vector comprising
the nucleic acid sequence encoding a polypeptide into a
second cell sample of E. coli, expressing the polypeptide
in the second cell sample and measuring the level of
20   polypeptide expression in the second cell sample. In one
embodiment this method includes the added step of selecting
cells from the first cell sample that exhibit optimized
polypeptide expression relative to the level of polypeptide
expression observed in the measuring step. Preferably, the
25   optimized polypeptide expression is a level of polypeptide
expression greater than or equal to the level of
polypeptide expression obtained from the measuring step. In
another preferred embodiment, the method additionally
comprises the step of determining the nucleic acid sequence
30   of the random oligonucleotide introduced into the position
in the vector, previously occupied by the signal sequence,
isolated from cells identified by the selecting step. In
yet another preferred embodiment of the invention the
concentration of ampicillin is at least 30 µg/ml and the

-16-

nucleic acid sequence encoding polypeptide is a sequence
encoding a single-chain antibody.

In another method of this invention, a method is
disclosed for identifying a novel nucleic acid sequence
5   encoding a protein trafficking signal that directs a
polypeptide to a desired location in a cell, comprising the
steps of: (a) identifying a region of nucleic acid
containing at least one protein trafficking signal to be
mutagenized; (b) preparing a nucleic acid vector comprising
10 the protein trafficking signal sequence to be mutagenized
and a nucleic acid sequence encoding at least one
polypeptide; (c) deleting the protein trafficking signal
sequence from the vector; (d) producing a pool of random
oligonucleotides; (e) using a polymerase chain reaction to
15 introduce at least one of the random oligonucleotides into
the position previously occupied by the protein trafficking
signal in a plurality of vectors to generate a pool of
mutagenized vectors; (f) introducing the mutagenized
vectors into a cell sample; (g) assaying for the location
20 of the polypeptide in the cell sample; (h) selecting cells
expressing the polypeptide in the desired cell location;
and (i) determining the nucleic acid sequence of the
protein trafficking signal contained in the mutagenized
vector located in the cells of step (h).

25   In a preferred embodiment of the trafficking
method of this invention, the protein trafficking sequence
to be mutagenized is located in the nucleic acid sequence
encoding the polypeptide and in another embodiment the
polymerase chain reaction is the enzymatic inverse
30 polymerase chain reaction and the nucleic acid sequence
encodes an antibiotic resistant gene and the selection step
consists essentially of growing the cell sample in the
presence of an antibiotic. Preferably the desired cell
location is extracellular.

SUBSTITUTE SHEET (RULE 26)

-17-

        In another aspect of this invention, a method is
provided for creating and isolating novel regulatory
sequences useful for optimizing the expression of a
recombinant polypeptide in prokaryotic cells comprising the
5   steps of: (a) identifying at least one regulatory region
within a nucleic acid sequence to be mutagenized; (b)
preparing a nucleic acid vector suitable for expressing
polypeptide in a prokaryotic cell, the nucleic acid vector
comprising the regulatory region and a nucleic acid
10 sequence encoding at least one polypeptide operably linked
to the regulatory region; (c) introducing the vector into a
first prokaryotic cell sample and expressing the
polypeptide encoded by the vector in the first cell sample;
(d) measuring the level of polypeptide expression in the
15 first prokaryotic cell sample; (e) deleting the regulatory
region from the vector; (f) producing a pool of random
oligonucleotides; (g) using a polymerase chain reaction to
introduce at least one random oligonucleotide into the
position previously occupied by the regulatory region in a
20 plurality of vectors to generate a pool of mutagenized
vectors; (h) introducing the mutagenized vectors into a
first prokaryotic cell sample; (i) assaying for the
expression of the polypeptide in the second prokaryotic
cell sample; (j) selecting cells exhibiting optimized
25 polypeptide expression relative to step (d); and (k)
determining the nucleic acid sequence of the mutagenized
regulatory region located within the vector introduced into
the cells of step (j).

        In a preferred embodiment of this method, the
30 regulatory region is selected from the group consisting of
a signal sequence, a ribosome binding site, a promoter
sequence, a translational regulatory sequence, a
transcription regulatory sequence and a protein trafficking
sequence. In one embodiment the nucleic acid sequence
35 encoding polypeptide encodes an antibiotic resistant gene

SUBSTITUTE SHEET (RULE 26)

-18-

and the selection step additionally consists of growing the
second prokaryotic cell sample in the presence of an
antibiotic. Preferably the first and second cell samples
are derived from the same cell type.

5   Figure 1 diagrams the assembly of the single
chain antibody β-lactamase fusion protein and the
incorporation of the fusion protein with the Omp A signal
sequence into the expression vector pCCHAsc1-ampRv2.

          Figure 2 illustrates the strategies followed to
10   delete the Omp A regulatory region, to incorporate the
random oligonucleotides, and to generate a library of
random signal peptide mutants.

          Figure 3 is a diagram of plasmid pCLA3ampR that
was used to reclone identified signal sequences to test for
15   signal sequence effectiveness.

          Figure 4 is a comparison of the hydrophobicity
and the alpha and beta region analysis of two novel signal
peptide sequences identified using the methods of this
invention as compared with known signal peptide sequences.

20   Figure 5 illustrates the construction of variable
length signal peptide libraries using the methods of this
invention.

          Figure 6 is a diagram of the plasmid
modifications used to produce the plasmid containing the
25   ribosome binding site library.

          Figure 7 is a diagram of plasmid pGCEMK.

          Figure 8 is a diagram of plasmid pNCEMG1.

          Figure 9 illustrates sequences of some of the
oligonucleotides containing the ribosome binding site and
30   signal peptide sequences identified using the methods of
this invention.

          This invention provides methods useful for
creating novel regulatory sequences that affect recombinant
protein expression within either a prokaryotic or
35   eukaryotic cell. In particular, these methods

SUBSTITUTE SHEET (RULE 26)

-19-

advantageously facilitate the optimization of recombinant
protein expression in a cell. Using the techniques of this
invention, modifications can be incorporated into more than
one regulatory region and the cumulative effect of these

5   mutations can be assessed by isolating clones expressing
the desired level of recombinant protein expression.

    The term "polypeptide" and "protein" are used
interchangeably throughout this text.

    The term "heterologous protein expression" is

10 used herein to refer to protein that is not native to the
host cell expressing the heterologous protein.

    The term "optimized protein expression" is used
herein to refer to the cumulative conditions that provide
an optimal level of protein expression for a particular

15 protein in a particular cell system. Under one set of
laboratory conditions, optimized protein expression may
refer to the highest available level of protein expression,
while in another set of laboratory conditions, optimized
protein expression may refer to low level protein

20 expression because for that particular application low
level protein expression is preferred. In yet another set
of laboratory conditions optimized protein expression may
refer to the level of protein expression that can coexist
with cell life in situations where, under standard

25 conditions, the protein would be cytocidal to the cells.

    The term "operably linked" is used to mean
regulatory regions that ultimately influence, or effect the
level of recombinant protein expression from a nucleic acid
vector.

30   The term "regulatory region is outside of a
nucleic acid sequence encoding polypeptide" is used herein
to indicate that the regulatory region of interest is not
located within the coding region. For this purpose, the
term coding region is used to denote that region of DNA

SUBSTITUTE SHEET (RULE 26)

-20-

that corresponds to RNA beginning with a methionine codon and ending with a polyadenylation signal.

While the methods of this invention are suitable for the optimization of recombinant protein sequences in
5   either prokaryotes or eukaryotes, it is contemplated that these methods are particularly useful for the optimization of heterologous protein sequences in either prokaryotes or eukaryotes. In particular, these methods are useful for optimizing the expression of eukaryotic protein in
10 prokaryotic cells.

Expression of eukaryotic protein in a prokaryotic system is often commercially desirable. Prokaryotic cells typically have shorter doubling times than eukaryotic cells and they are easier and less expensive to grow in large
15 quantity than eukaryotes. However, optimizing expression of a eukaryotic protein in a prokaryotic cell such as *E. coli* has heretofore been an inefficient process. Typically a eukaryotic gene sequence encoding protein that is incorporated into a prokaryotic expression vector is
20 expressed at relatively low levels or at levels that result in prokaryotic cell death. Given the multitude of differences in protein expression between prokaryotes and eukaryotes, as outlined in the background of this invention, it is often difficult to identify exactly why a
25 particular eukaryotic sequence encoding protein is either not efficiently expressed or is bacteriocidal. The methods of this invention do not require a systematic assessment of each individual factor that can effect protein synthesis. Rather, the methods of this invention employ a novel random
30 mutagenesis and selection strategy that advantageously permits one of skill in the art to identify clones that yield optimized levels of heterologous protein expression by assessing the cumulative effect of all of the changes incorporated into the expression vector. Thus, by the
35 method of this invention, it is possible to select clones

SUBSTITUTE SHEET (RULE 26)

-21-

having mutagenized regulatory sequences individually
tailored to optimize expression of any given protein in any
given expression cell.

Most mutagenesis techniques involve the
5   systematic replacement of one or more nucleotides with
other known nucleotides. A variety of site-directed
mutagenesis strategies are known in the art and are
commercially available in kit format (see BioRad, Richmond,
CA; Stratagene, La Jolla, CA; or Invitrogen, San Diego,
10   CA.) Here, using site-directed mutagenesis, one begins with
a known, efficient regulatory region and improves it by
introducing single or multiple point mutations. In this
procedure, the mutagenized regulatory region may or may not
be biased according to what is known about the chemical
15   properties of the wild-type regulatory region. For site-
directed mutagenesis techniques used to optimize a
prokaryotic signal peptide through the modification of a
hydrophobic region of the signal peptide see Kendall, et
al. *Nature* 321:706-708, 1986. For a review of site-
20   directed mutagenesis strategies for prokaryotic signal
peptides see Ryan, et al. *J. Biol. Chem.* 261:3389-3395,
1986; Park, et al. *Agric. Biol. Chem.* 55:1745-1750, 1991;
vigal, et al. *Mol. Gen. Genet.* 231:88-96, 1991; Borchert,
et al. *J. Bacteriol.* 173:276-282, 1991; Morioka-Fujimoto,
25   et al. *J. Biol. Chem.* 266:1728-1732, 1991. Similarly,
cassette mutagenesis strategies are also known in the art.
These strategies involve removing a section of DNA between
two restriction sites and replacing it with a different DNA
sequence that is bounded by the same restriction sites
30   (Richards, J.H., (1991) "Cassette Mutagenesis" in <u>Directed
Mutagenesis, A Practical Approach</u>, M.J. McPherson ed. IRL
Press). Chou, et al. (*J. Biol. Chem.* 265:2873-2880, 1990)
studied the hydrophobicity requirements of the prokaryotic
signal sequence using cassette mutagenesis. Ngsee, et al.
35   used cassette mutagenesis in eukaryotic cells to study the

SUBSTITUTE SHEET (RULE 26)

-22-

effect of signal sequence mutants on protein expression
(*Mol. Cell. Biol.* 9:3400-3410, 1989). These techniques
replace identified sequences with known alternatives.
However, unlike the mutagenesis strategies of this
5   invention, these traditional site-directed and cassette
mutagenesis techniques are not efficient methods for
creating novel functional sequences.

In the practice of this invention, random or
biased mutagenesis techniques are useful methods for
10   generating a pool of novel mutants. Regions of
oligonucleotides, or whole oligonucleotides, are
synthesized using methods that permit the random
incorporation of nucleotides. Random mutagenesis combined
with good selection systems can be used to identify
15   functional regulatory sequences lacking any predisposed
bias to previously identified sequences. The collection of
random sequences is incorporated into the gene sequence in
place of the native sequence.

For purposes of this application, the term random
20   mutagenesis will be used to encompass biased mutagenesis
techniques. Biased mutagenesis is a type of random
mutagenesis in which pools of oligonucleotides are prepared
to accommodate the incorporation of more than one type of
nucleotide in a given location within a nucleic acid
25   sequence. Biased mutagenesis techniques are particularly
useful for generating pools of mutants that maintain the
consensus patterns of charge, hydrophobicity,
hydrophilicity, etc. within a particular region of protein
encoded by a nucleic acid sequence. In this way the
30   chemical characteristics of a molecule can be maintained
while the actual nucleic acid sequence and amino acid
sequence of an encoding protein is varied. The application
of biased mutagenesis to regulatory regions has the
advantage that optimized sequences can be identified that
35   are chemically biased toward native sequences but reflect a

SUBSTITUTE SHEET (RULE 26)

-23-

wider variation than what has previously been identified in
the art. Some of these sequences may direct protein
expression more efficiently than other sequences heretofore
identified.

5   Strategies of random mutagenesis to select clones
containing functional mutations is complicated by the
contaminating background level of clones containing
nonmutated sequences. The presence of nonmutated sequences
complicates selection and screening strategies by

10   decreasing the variability within a library and by
increasing the number of false positive clones. A large
number of the clones produced by traditional random
mutagenesis techniques are false positive because they
contain the functional nonmutated sequence. Since the

15   presence of novel optimized sequences from a random or
biased pool of nucleotides is a relatively rare event,
contaminating protein expression directed from nonmutated
sequences can easily block the signal of the rarer,
optimized sequence.

20   In this invention the drawback of non-mutated
clones is avoided by substantially removing the wild-type
sequence and replacing it with the randomly mutagenized
sequence before selection of clones for function
characteristics.

25   As noted, this invention is particularly useful
for generating novel regulatory sequences associated with
protein expression. The eukaryotic regulatory sequences
contemplated within the scope of this invention include,
but are not limited to promoters, upstream promoter

30   elements, enhancers, transcriptional regulatory elements,
translational regulatory proteins such as translation
repressor proteins, signal sequences, protein targeting
sequences, chaperon proteins (these binding proteins coat
polypeptide chains as they emerge from ribosomes or

35   membranes, preventing aggregation reactions and premature

SUBSTITUTE SHEET (RULE 26)

-24-

folding (Flynn, et al. *Nature* 353: 726, 1991) and sequences
directing post-translational modifications and the like.
The prokaryotic regulatory sequences contemplated within
the scope of this invention include, but are not limited

5   to, promoters, transcription regulatory sequences,
translation regulatory sequences including ribosome binding
sites, sequences associated with attenuation, protein
targeting sequences, signal sequences, and the like.
        Thus, the regulatory sequences may be in any

10   number of locations relative to the gene sequence encoding
protein, for instance, within the translated region of a
coding sequence, within a transcribed region of a coding
sequence, or external to the coding sequence. For example,
the signal sequence for both eukaryotic and prokaryotic

15   cells is positioned at the amino terminus of the nucleic
acid sequence encoding protein. Since the primary sequence
of the signal peptide has a direct effect on the
translocation of the associated protein (Inouye, et al.
Proc. Natl. Acad. Sci., USA 74: 1004-1008, 1977), the

20   signal peptide is a candidate regulatory sequence suitable
for the methods of this invention. Example 1 illustrates
the application of the methods of this invention to the
generation of novel signal sequences. As a second example
to illustrate that the methods of this invention can be

25   applied to a variety of regulatory sequences, Example 3
provides an exemplary strategy useful for generating novel
ribosome binding sites.
        The regulatory sequences created by the methods
of this invention may be of any length. For example, while

30   most signal sequences fall between a range of between 18-24
amino acids, it is contemplated that novel optimized signal
sequences could be created in a variety of lengths. Example
2 provides methods for generating libraries of signal
peptides having lengths of 12, 16 or 20 amino acids.

35   Similarly these methods could be applied to other

SUBSTITUTE SHEET (RULE 26)

-25-

regulatory regions such as ribosome binding sites, protein
targeting sequences and the like.

Identification of a Regulatory Region to be Mutagenized

5   As a first step for practicing the methods of
this invention, the position of the regulatory region to be
mutagenized is identified. Regulatory regions contemplated
within the scope of this invention have been discussed
above. The relative position of the regulatory region can
10   be identified within a particular nucleic acid sequence by
studying the consensus data for that particular regulatory
region--available from the prior art. There exists in the
literature a wealth of information concerning the consensus
positions of known regulatory regions. For example,
15   regions encoding signal peptides have been localized for
both prokaryotic and eukaryotic genes as have regions
encoded on mRNA which are required for ribosome binding
(Protein Targeting. supra, pp. 52-61; Kozak, *Microbiol.*
Rev. 47:1-45, 1983; and Kozak, *Nucl. Acids Res.* 15: 8125-
20   8147, 1987). The precise identification of the position of
a regulatory region within or in association with a nucleic
acid sequence encoding protein is not necessary. Rather,
any particular region of the nucleic acid containing a
putative regulatory region postulated to exist within a
25   nucleic acid sequence can be subjected to the mutagenesis
techniques of this invention. These putative regulatory
regions may be postulated in the art or they may be
identified by trial and error using site-directed
mutagenesis, or other mutagenesis techniques known in the
30   art.

Preparing the Nucleic Acid Vector

            Once the regulatory region to be mutagenized is
identified, it is then incorporated into a nucleic acid
35   vector together with a nucleic acid sequence encoding at

-26-

least one protein under the control of the regulatory
region to be mutagenized. Any suitable expression vector
may be used in this invention and exemplary vectors are
provided in the Examples below. Those with skill in the
5   art will appreciate that the choice of vector is limited to
those vectors capable of directing expression of the
nucleic acid sequence encoding protein and to those vectors
that can incorporate and support the function of the
regulatory region to be mutagenized. Further, the choice
10 of vector is limited by the cell type selected; not all
vectors and not all regulatory elements necessary for
recombinant protein expression function in all cell types.
As a general rule, eukaryotic expression vectors are
suitable for protein expression in eukaryotes and
15 prokaryotic expression vectors are suitable for
prokaryotes. Both types of vectors are commercially
available and those with skill in the art of molecular
biology will be able to select the appropriate vectors
suitable for recombinant protein expression within a given
20 cell type. In addition, the following vectors used in the
Examples of this invention (pCLA3ampR; pCCHAsc1-ampRV2; and
pCampR) have been deposited with the American Type Culture
Collection (Bethesda, Maryland) under the Budapest Treaty.
        Methods for incorporating a particular region of
25 nucleic acid into a nucleic acid vector are well known in
the art of molecular biology (see Sambrook, et al., supra).
For example, short regions of nucleic acid (less than 200
bp) can be prepared by generating overlapping
oligonucleotide fragments complementary to the nucleic
30 acid. These oligonucleotides are then hybridized to one
another, ligated and incorporated into an appropriate
expression vector. Alternatively, those with skill in the
art of molecular biology will be able to use the polymerase
chain reaction technology to amplify a suitable nucleic
35 acid fragment containing the regulatory gene of interest

SUBSTITUTE SHEET (RULE 26)

-27-

and incorporate this fragment into the expression vector of
choice (see generally Erlich, H.A. PCR Technology:
Principles and Applications for DNA Amplification, 1992.
W.H. Freeman and Co., New York).

5   As noted in the preceding paragraphs, the nucleic
acid vector additionally contains a nucleic acid sequence
encoding a recombinant protein product such that the
regulatory region is operably linked to and/or regulates
expression of the protein encoded by the gene. Since the
10   regulatory region may be positioned in the nucleic acid
vector as part of the nucleic acid sequence encoding
protein or as a nucleic acid sequence adjacent to the
nucleic acid sequence encoding protein, the assembly of the
nucleic acid vector containing the regulatory region to be
15   mutagenized may require multiple steps. Those skilled in
the art of molecular biology will be able to select an
initial nucleic acid vector suitable for expressing a
protein product from a gene sequence in a host cell and to
incorporate both the gene sequence encoding the protein
20   product and the gene sequence encoding the regulatory
region into the vector using techniques of genetic
engineering well known in the art. Incorporation of both
the regulatory sequence and the gene sequence encoding the
protein product into the nucleic acid vector is designed
25   such that once the nucleic acid vector is introduced into a
compatible cell sample, if the regulatory sequence is
functional, protein expression will be detected. Compatible
cells are those cells capable of expressing a protein from
the nucleic acid vector when the vector contains those
30   regulatory elements that facilitate protein expression
(i.e. origins of replication, etc.).

              It is anticipated that other gene sequences may
additionally be incorporated into the nucleic acid vector
of this invention. For example, gene sequences coding for
35   antibiotic resistance, or other gene sequences that

SUBSTITUTE SHEET (RULE 26)

-28-

facilitate cell selection may be introduced into the
vector. Similarly, the nucleic acid sequence encoding
protein can itself encode an antibiotic resistance gene or
a portion of an antibiotic resistance gene. Other
5   sequences that may be incorporated into the vector include
but are not limited to promoters, enhancers, polyA+
adenylation sites, origins of replication (eukaryotic
and/or prokaryotic), and specific restriction endonuclease
cleavage sites (or arrays of sites). Exemplary nucleic
10   acid vectors containing a regulatory region to be
mutagenized and a nucleic acid sequence encoding protein
are provided in Examples 1-3 as well as Figures 1-3 and 6.

Expression of the Recombinant Protein under Control of the
15   **Regulatory Region to be Mutagenized**
         In order to identify enhanced or optimized
protein expression from a particular mutagenized regulatory
region relative to the regulatory region to be mutagenized,
some base line information should be obtained for protein
20   produced under the control of the nonmutagenized regulatory
region. This data is later compared to the level of
protein expression influenced by the mutated regulatory
region. This data may be obtained from the literature, or
the data may be obtained first hand. As one method for
25   obtaining the data first hand, the vector containing the
nonmutagenized regulatory region is introduced into a cell
sample that is able to support the expression of the gene
sequence under the control of the regulatory region.
         There are a variety of methods recognized in the
30   art for introducing nucleic acid vectors into prokaryotic
and eukaryotic cells. Depending on whether the cells are
prokaryotic or eukaryotic, the vectors may be introduced as
viral vectors, electroporated into the cells, bacterial
cell walls may be permeabilized, the vectors may be
35   introduced through standard precipitation methods or

SUBSTITUTE SHEET (RULE 26)

-29-

through the use of lipophilic agents. Such methods are
commercially available as kits and detailed methods are
readily available in the literature. Methods for
transforming *E. coli* with nucleic acid vectors are provided
5   in Example 1.
        Next, the level of protein expression from the
nonmutagenized nucleic acid vector is measured in the cell
sample. Proteins of interest may be concentrated from cell
supernatants or cell lysates and quantitated by
10 chromatographic methods, enzyme-linked immunoadsorbant
assays (ELISA), western blot assays, radio-immunoassays,
gel electrophoresis, or the like. Exemplary assays for
measuring the amount of protein expressed in the
periplasmic space of *E. coli* are provided in Example 1 and
15 are detailed in Stemmer, et al. *BioTechniques* 13:114-220,
1992, hereby incorporated by reference and U.S. Patent
Application No. 07/641,140, filed April 26, 1991, and U.S.
Patent Application No. 07/806,154, filed December 12, 1991.
Alternatively, an exemplary assay for measuring the amount
20 of protein expressed in a eukaryotic cell is provided in
Example 4. Methods to detect protein expression from a
nucleic acid vector in a cell will vary depending on the
gene sequence encoding protein. Those with skill in the
art of molecular biology are well versed at selecting gene
25 sequences encoding protein and matching these sequences
with methods suitable for detecting protein expression
derived from that gene sequence in a particular cell type.
The value of protein expression obtained from cells
expressing recombinant protein under the influence of the
30 nonmutagenized regulatory region provides a level of
expression that can be compared with that of cells
expressing protein from vectors with mutagenized regulatory
regions.

-30-

Deletion of the Regulatory Region to be Mutagenized

An important aspect of this invention involves the deletion of the region to be mutagenized before the pool of random oligonucleotides are incorporated into the

5   vector. This step of the invention advantageously removes the endogenous background level of protein expression that contaminates mutagenic libraries produced by other methods. The regulatory region can be deleted from the nucleic acid vector in a variety of ways. For example, during vector

10 construction, two identical vectors may be simultaneously assembled such that one contains and the other lacks the mutagenized regulatory region. The regulatory region can be deleted from the vector using PCR technology and an exemplary strategy is disclosed in Example 1. Similarly

15 the regulatory region can be deleted through the use of restriction endonuclease recognition sites bracketing the regulatory region.

Preparation of Random Oligonucleotides

20 As a next step for practicing the methods of this invention, a pool of random oligonucleotides is prepared that are suitable for PCR methods and can be readily incorporated into the position in the vector previously occupied by the regulatory region. These oligonucleotides

25 are positioned in the nucleic acid vector in place of the deleted regulatory region. As noted previously, the term "random oligonucleotides" includes biased oligonucleotides. Methods for generating random oligonucleotides on a DNA synthesizer are well known in the art. In this invention

30 the random oligonucleotides are of a size and design useful for directing a polymerase chain reaction and for incorporating the nucleic acid vector containing the nucleic acid sequence encoding protein. That is, the oligonucleotides should contain regions of homology to the

35 vector, they should include a restriction endonuclease site

SUBSTITUTE SHEET (RULE 26)

-31-

to facilitate recircularization of the plasmid following
incorporation of the vector, and they should include a
region containing the random or biased nucleotides. The
oligonucleotides can be of any desired length, and the
5   length of the oligonucleotide will depend on the type of
regulatory sequence and the number of random
oligonucleotides that will be incorporated into the vector.
      The oligonucleotides prepared in this invention
contain at least one region that incorporates more than one
10   nucleotide at a given position within the oligonucleotide.
This pool of random oligonucleotides is incorporated into
the vector to produce a library of clones containing random
regulatory sequences. Libraries prepared from random
oligonucleotides, and particularly libraries prepared using
15   the methods of this invention wherein the regulatory region
to be mutagenized is deleted, have the advantage that the
selected active clones will likely be novel regulatory
sequences. Because of the large theoretical number of
random regulatory regions be they signal peptides, ribosome
20   binding sites, or the like, the methods of this invention
are more likely to generate novel regulatory sequences that
direct increased or optimized protein expression than other
methods currently available in the art.
      One potential difficulty of a completely random
25   regulatory library is that the majority of the sequences
generated are inactive. This means that the libraries must
be sufficiently large in order to identify a sufficient
number of active clones for further study. If the goal of
the mutagenesis procedure is to generate a regulatory
30   region that significantly increases the expression of the
recombinant protein, then this type of regulatory region
would occur infrequently in a random library. For example,
in an *E. coli* library, if the probability of finding a
"super sequence" is 10^{-9} there is a good chance that it
35   would not appear in a library of 10^{8} clones.

-32-

Exemplary Calculation to Estimate the Library Size
Necessary for Identifying Novel Regulatory Sequences

      The following example is provided as one method
for estimating an adequate library size that would be
5   required to identify a novel, optimized sequence. This
example is specifically addressed to the library size
required to identify an optimized signal peptide in a
prokaryotic cell using biased oligonucleotides. It is
contemplated that those with skill in the art would be able
10   to apply the strategy exemplified by the calculations used
in this example to determine the library size required for
other regulatory regions in either prokaryotes or
eukaryotes.

      The known requirements for a signal sequence are
15   first incorporated into the calculation. Known signal
peptides contain one or two positively charged amino acids
near the N-terminus, a stretch of 8-12 hydrophobic amino
acids and a small amino acid such as alanine, glycine or
valine positioned at the C-terminus. For consensus regions
20   in prokaryotic signal sequences see MacIntyre, et al. *Mol.*
Gen. Genet. 221:466-474, 1990; Gennity, et al. *J.*
Bioenergetics and Biomembranes 22:233-269, 1990; Laforet,
et al. *J. Biol. Chem.* 266:1326-1334, 1991 and Yamane, et
al. *J. Biol. Chem.* 263:19690-19696, 1988 and for consensus
25   regions and exemplary eukaryotic signal sequences see Von
Heijne, et al. *Eur. J. Biochem.* 133:17-21, 1983.

      The fraction of active leader peptides in this
example is (f) = p1*, p2*, p3* ....pn, where p1, p2, p3...pn
represent the probability of a given amino acid or amino
30   acids, occurring in a given position. In a hypothetical
library containing a 22 amino acid signal peptide, the
peptide can be diagrammed as [M][+][+][n]$_{19}$[S] where each []
represents an amino acid, [M] represents methionine, [+]
represents a positively charged amino acid, [n] represents
35   any hydrophobic amino acid and [S] represents a tiny amino

SUBSTITUTE SHEET (RULE 26)

-33-

acid. The fraction of active leader peptides (f) =
$(1)(.15)(.15)(.66)^{18}(.2) = 7.3 \times 10^{-6}$ Pugsley, supra, p. 47,
Sambrook, supra, Appendix p. 6. Thus a signal peptide of
this particular configuration would occur once in 300,000

5   random clones. This value reflects the incidence of an
active signal peptide within a library and not the
frequency of an optimized signal peptide. One would expect
that the frequency of an optimized signal peptide occurring
in a random library to be less; therefore the library size

10   would need to be larger.
          Several assumptions were made in the above
calculation. First, it was assumed that two positive
charges are required at the N-terminus. In fact, there are
known signal peptides that carry only a single positive

15   charge at their N-terminus. If only a single positive
charge is required, active leader peptides will occur at a
higher frequency. Similarly, for this calculation, the
hydrophobic region was limited to only hydrophobic amino
acids. There are some cases where amino acids such as gln

20   or asn or even acidic or basic residues occur in this
region. Thus, the calculation of $(0.66)^{18}$ may be too
restrictive. If 2 amino acids in the hydrophobic stretch
are permitted to be any of the 20 amino acids, then we
arrive at a value of $(0.66)^{16}$. Finally, the impact of the

25   degeneracy of the code was ignored in these calculations.
Although the last amino acids in known signal peptides are
usually glycine or alanine, these amino acids are generally
believed to be required for signal peptide cleavage. Since
signal peptide cleavage is not required for translocation

30   and β-lactamase activity, that requirement can be
eliminated. If these less restrictive requirements are
used, the frequency of a leader peptide occurring in a
random library is 0.02% or 1 in 5000 random sequences.
These types of calculation are important for estimating the

-34-

required library size needed to identify a novel, optimized sequence.

In practical terms, a biased library is difficult to construct due to the constraints on the usage of the genetic code. For example, if one tries to use oligonucleotide doping to produce either a lysine or arginine in a given position, the codon for glutamine will also occur. Similarly, due to the degeneracy of the genetic code, it is impossible to constrain an amino acid to solely a hydrophoibic amino acid using single base oligonucleotide synthesis from variable base mixtures. In some cases it is possible to use doping to achieve the desired amino acid mixture. For example, a 50% occurrence of alanine or glycine will occur if the sequence 5' G(G/C)X is used. The successive addition of pre-made trimers can obviate this problem. Precise ratios of desired amino acids can be designed into the library through use of premixed trinucleotide codon triplets during DNA synthesis (Sondek, et al. *Proc. Natl. Acad. Sci.* (USA) 89:3581-3585, 1985). Given current DNA synthesizer technology this would be a time consuming and expensive undertaking.

As another method for generating a biased library the oligonucleotide column swapping DNA synthesis technique developed by Glaser, et al. is used (*J. Immunol.* 149:3903-3913). This method produces a 50% chance of a mutation at any desired amino acid. This is a codon based mutagenesis approach that allows the formation of large diverse libraries. The column-swapping manipulation causes one-half of each synthesis step to contain random codons, while the other half contains wild-type codons, at each position to be mutated. In this way, one can mutate large stretches of DNA without a high frequency of stop codons, as only one of the three possible stop codons can occur, and this only at a frequency of 1 in 32. The use of a suppressor tRNA for this stop codon can reduce the frequency to zero.

SUBSTITUTE SHEET (RULE 26)

-35-

However, the preparation of biased
oligonucleotides to identify optimized regulatory sequences
is not an absolute requirement for the methods of this
invention. Optimized sequences can be identified from
5   clones incorporating random oligonucleotides containing no
chemical bias. In a preferred embodiment, illustrated in
Example 1, the invention demonstrates that even with no
preconceived notions about the specific requirements for a
regulatory region, the bacterial host will naturally select
10   those regulatory regions that it finds functional in
expressing the protein of interest.

**Incorporation of the Oligonucleotides into the Nucleic Acid
Vector**
15   As a next step for practicing the methods of this
invention, the mutagenized oligonucleotides are introduced
using a polymerase chain reaction into the position in the
vector previously occupied by the deleted regulatory
region. There are a variety of polymerase chain reaction
20   methods known in the art that are suitable to the methods
of this invention. The examples below use the enzymatic
inverse polymerase chain reaction (see Stemmer, et al.
BioTechniques 13:114-220, 1992, hereby incorporated by
reference and U.S. Patent Application No. 07/641,140, filed
25   April 26, 1991, and U.S. Patent Application No. 07/806,154,
filed December 12, 1991) as a preferred PCR strategy.
EIPCR advantageously allows targeted mutagenesis of a given
DNA sequence and rapid selection of active clones from a
large library. Other PCR techniques suitable for use in
30   the methods of this invention include, but are not limited
to, inverse PCR (Hemsley, et al. *Nucl. Acids Res.* 17:6545-
6551, 1989), recombinant circle PCR mutagenesis (Jones, et
al. *BioTechniques* 8:178-183, 1990), recombination PCR
(Jones, et al., *BioTechniques* 10:62-66, 1991) and standard

-36-

PCR techniques such as those described by Saiki, et al.
(*Science* 239:487-491, 1988).

Briefly, the enzymatic inverse polymerase chain
reaction (EIPCR) involves the use of oligonucleotides
5   containing a class IIS restriction endonuclease recognition
site and the associated downstream class IIS cleavage site.
The EIPCR technique uses a small circular vector and the
PCR primers are designed to overlap at their class IIS
cleavage sites. In one embodiment, one oligonucleotide of
10   the oligonucleotide pair used in the EIPCR reaction
contains in a first 5' to 3' orientation a 5' tail, a class
IIS restriction endonuclease recognition site, a class IIS
cleavage site, the region containing the mutation and a
region complementary to the nucleic acid vector. The other
15   oligonucleotide of the oligonucleotide pair comprises in a
second 5' to 3' orientation; a 5' tail, a class IIS
restriction endonuclease recognition site, a class IIS
cleavage site and a region complementary to the nucleic
acid vector. In another preferred embodiment, both
20   oligonucleotides of the oligonucleotide pair used in the
EIPCR reaction contain the first 5' to 3' orientation.
Further details of primer design applicable to the
mutagenesis application of this invention are provided in
Examples 1-3.
25   The amplified linear vectors are digested with
the appropriate restriction endonuclease and recircularized
to produce a pool of mutagenized vectors. The pool of
mutagenized vectors is purified and aliquots are introduced
into a cell sample using any of the methods disclosed in
30   association with the discussion relating to cell
transformation or transfection discussed above. Example 1
uses electroporation to introduce the mutagenized vectors
into the cell sample. Preferably the cell samples used for
both the mutagenized and nonmutagenized vector
35   transformations or transfections are derived from the same

-37-

cell type. In this way, protein expression from cells
containing the nucleic acid vector having the
nonmutagenized regulatory region can be directly compared
to expression from cells containing the nucleic acid vector
5   with the mutagenized regulatory region.

**Assaying for the Expression of the Recombinant Protein from
the Mutagenized Vectors**

Following introduction of the mutagenized vectors
10 into the cell sample, the cell sample is assayed for the
expression of the recombinant protein, that is, the protein
associated with the mutagenized regulatory region. As noted
in the section entitled "Expression of the Recombinant
Protein under Control of the Regulatory Region to be
15 Mutagenized", the particular assay for assessing the level
of protein expression will depend on the particular protein
being expressed as well as on the cell type. Exemplary
methods for assessing protein expression in the periplasmic
space of prokaryotes are provided in Example 1. For
20 enhanced protein expression, cells are selected that
exhibit protein expression from the nucleic acid vector at
a level greater than or equal to the level of protein
expression observed in the cell sample containing
nonmutagenized vector. For optimized protein expression,
25 other characteristics, in addition to the level of protein
synthesis, may be important for clone selection. These
include cell growth characteristics, the location of
protein expression or the like.

Referring to Example 1 below, the processes of
30 transcription, translation, translocation and signal
peptide cleavage are parameters that are affected when the
signal sequence is mutated. The use of a random signal
sequence allows for selection of clones that are optimized
for all of these parameters using the internal mechanisms
35 of the host cell. For example, a random signal sequence

SUBSTITUTE SHEET (RULE 26)

-38-

that generates a nonfunctional secondary structure in RNA
would preclude efficient translation and positive clones
containing this sequence would not be identified. Assays
for recombinant protein expression from the mutagenized

5   vectors permits the cumulative effect of the mutated
sequence on protein expression to be quantitated. In
addition, these parameters are optimized with respect to
the individual requirements of the product proteins being
expressed by the regulatory regions such as glycosylation

10   requirements, folding requirements for functionality, and
the like.

As a final step in the methods of this invention,
the nucleic acid vector is isolated from cells derived from
the cell sample that contain mutagenized vector and exhibit
15   enhanced or optimized expression of the recombinant
protein. The purified vector is subjected to DNA sequence
analysis to determine the nucleic acid sequence of the
mutagenized regulatory region corresponding to the random
oligonucleotide. DNA sequencing techniques are well known
20   in the field of molecular biology, thus no further
disclosure of sequencing techniques is required.

Identification of optimized regulatory sequences
is a particularly important step in the method of
optimizing protein expression from any number of vectors.
25   Thus, this invention provides a method for both optimizing
regulatory regions associated with protein expression and
for identifying the optimized sequences of these regions.
Those with skill in the art of molecular biology will be
able to incorporate these optimized sequences into other
30   constructs thereby optimizing expression of any number of
proteins from a variety of vectors introduced into a
variety of cell types.

-39-

Incorporation of Selectable Markers into the Nucleic Acid Vector

Rather than improving upon existing regulatory regions, this invention creates novel regulatory sequences,
5   including novel signal peptides. As noted in the section entitled "Exemplary Calculation to Estimate the Library Size Necessary for Identifying Novel Regulatory Sequences", one would expect that a library of 21 random amino acid sequences would contain a relatively low percentage of
10 functional secreting clones. However, by attaching the random regulatory region to a selectable marker, particularly a marker which requires secretion for activity, the library can be restricted to only clones containing functional signal peptides. Selectable marker
15 proteins, particularly those in prokaryotes require transport to the periplasmic space to be active, have the advantage that when a regulatory region is operably linked to the amino-terminus of a heterologous protein containing the C-terminus of a periplasmic protein, this will result
20 in transport to the periplasmic space and the selectable marker is able to confer a subsequent advantage to cell growth under certain conditions. The in-frame fusion of a heterologous protein to a selectable marker provides a rapid means to select only active regulatory sequences.
25 Furthermore, if the selectable marker is an antibiotic resistant gene, then the use of increasing concentrations of antibiotic allows for selection of clones containing the most efficient regulatory sequences.

The enzyme β-lactamase represents one such
30 selectable marker because it is active only if translocated from the cytoplasm to the periplasm (Pluckthun, et al. *J. Biol. Chem.* 262(9): 3951-3957, 1987 and Francisco, et al. supra). While Example 1 uses a regulatory region operably linked to a eukaryotic protein fused to the C-terminus of
35 β-lactamase, it also is contemplated that the regulatory

SUBSTITUTE SHEET (RULE 26)

-40-

region can be operably linked to a full-length selectable marker such as β-lactamase (see Example 3). However, the creation of novel regulatory regions by assessing protein expression in the periplasmic space is not limited to

5   β-lactamase fusion proteins. Other selectable marker proteins can also be used. As an alternative, aminoglycoside 3'-phosphotransferase II can be used as a selectable marker in the same way. This enzyme is present in bacteria that are resistant to streptomycin, neomycin or

10 kanamycin. The enzyme renders the antibiotic inactive as it enters the cell through the periplasm. Other periplasmic enzymes include Streptomycin adenylating enzyme and alkyl sulfhydrolase.

        Other selectable markers contemplated for use in

15 the methods of this invention include binding proteins for various carbon sources. Proper secretion of this class of proteins leads to their presence on the outer membrane or periplasm, enabling the cell to internalize a particular sugar or amino acid. For example, maltose binding protein,

20 when properly secreted, allows bacteria to use maltose as the sole carbon source. If the maltose binding protein is not secreted, cells cannot grow on media containing maltose as the sole carbon source. In some cases, it is necessary to use the appropriate auxotrophic strain of E. coli. For

25 example a his-negative strain requires histidine in the media. Other selectable proteins that may be useful as fusion proteins include but are not limited to arabinose binding protein, arginine binding protein, cystine and diaminopimelic binding protein, galactose-glucose binding

30 protein, glutamate-aspartate binding protein, glutamine binding protein, histidine binding protein, leucine specific binding protein, lysine-arginine-ornithine binding protein, phosphate binding protein, ribose binding protein, sulfate binding protein, thiamine binding protein and

35 xylose binding protein.

SUBSTITUTE SHEET (RULE 26)

-41-

One can also use nonselectable markers. These
are proteins that do not necessarily give the cell any
physiological advantage. Their presence in the periplasm
is detected based on their interaction with specific
5   ligands. With the use of these markers it is possible to.
detect the periplasmic protein by probing with the
appropriate ligand. Examples of these markers include but
are not limited to hapten binding antibodies, cytochrome c,
nitrite reductase, phosphoglucomutase and phosphoglucose
10   isomerase. Therefore, while Example 1 employs a random
sequence library linked to a fusion protein of a single
chain antibody with an antibody resistance gene, the
regulatory region could also be linked solely to a sequence
encoding an antibody sequence and positive clones selected
15   by screening with labelled antigen. For identification of
intact colonies expressing antibody sequences in the
periplasmic space, the antigen is preferably a small hapten
capable of diffusing into the periplasmic space.

As noted above, any regulatory region affecting
20   the level of protein synthesis is contemplated as a
potential target for the optimization strategies of this
invention. Two specific prokaryotic examples are provided
below. These examples illustrate the methods of this
invention as they apply to optimizing signal sequences
25   (Example 1) and ribosome binding sites (Example 3). The
methods of this invention readily apply to mutagenesis
strategies directed to both these and other regulatory
regions in prokaryotic and eukaryotic cells.

As a particularly preferred strategy for
30   generating optimized signal sequences encoding signal
peptide, Example 1 details a method for creating novel
signal sequences for the expression of eukaryotic protein
in *E. coli*. In addition to creating novel signal
sequences, this method is also useful for optimizing the

SUBSTITUTE SHEET (RULE 26)

-42-

expression of a particular protein from a given construct in a given cell type.

The strategy detailed in Example 1 combines the mutagenesis strategies of this invention with antibiotic selection to improve secretion of Fv antibody molecules in *E. coli*. A fusion protein consisting of an N-terminal, hapten-binding single chain antibody (CHA 255) and the C-terminal of β-lactamase was constructed and expressed in active form in the periplasmic space of *E. coli* using the Omp A signal peptide. The fusion of β-lactamase with a second protein to study protein transport was reported by 10 Broome-Smith, et al. *Mol. Micro.* 4(10) 1637-1644, 1990. For studies linking Omp A with a prokaryotic protein as a fusion protein to β-lactamase see Francisco, et al. *Proc. Natl. Acad. Sci. USA* 89:2713-2717, 1992 and Meyer et al. disclosing a method for combining an Fab' antibody conjugate to β-lactamase, *Bioconj. Chem.* 3:42-48, 1992). For site-directed mutagenesis of a eukaryotic signal sequence associated with a eukaryotic protein fused to the C-terminus of β-lactamase for prokaryotic expression see Stahl, et al. (*Gene* 71:147-156, 1988)

Unlike previous mutagenesis strategies, the regulatory sequence of interest, here the Omp A signal peptide, was deleted and enzymatic inverse PCR was used to construct a library containing random signal peptides at the N-terminus of the fusion protein. None of the references cited above report the deletion of the regulatory region to be mutagenized followed by the isolation of vector containing that deletion to create large libraries of regulatory mutants uncontaminated with false positive clones. Functional signal peptides were selected by plating the clones on media containing ampicillin. Clones were identified that had β-lactamase activity comparable to or greater than that of the fusion protein containing the Omp A signal peptide. To ensure

SUBSTITUTE SHEET (RULE 26)

-43-

retention of antibody activity, colony filter lifts were
screened with a radioactively labeled hapten, ^{111}In-EOTUBE,
recognized by the CHA 255 antibody.

Example 1 provides a detailed method for
5   generating novel signal sequences that are 22 amino acids
in length. However, optimized signal sequences, like
optimized regulatory regions can be selected based on the
lengths of known regulatory regions, or, alternatively,
regulatory regions of any desired length can be selected
10   for optimization. Therefore, steps similar to those
outlined in Example 1 can be performed to generate
optimized signal sequences, or other regulatory regions of
any length. Example 2 provides methods for creating random
leader peptides of 12, 16, or 20 amino acids in length.

15   It is further contemplated that the methods of
this invention can be used to create novel regulatory
sequences for expression of either heterologous or
homologous protein. For example, Example 3 uses a native
E. coli protein to identify novel ribosome binding site
20   sequences. Once optimized ribosome binding site sequences
are identified they can be incorporated into other vectors
to study the effect of these sequences on heterologous
protein expression. In this example the steps outlined
above are followed for creating a library of random
25   regulatory sequences operably linked to protein except that
here the initial nucleic acid vector construct includes the
Omp A ribosome binding site linked to the Omp A signal
peptide located at the amino terminus of the β-lactamase
protein. The Omp A ribosome binding site sequence was
30   deleted by PCR and the resulting vector was used to
incorporate a library of random oligonucleotides. These
mutagenized vectors were introduced into E. coli and the
level of protein expression from the resulting clones was
assessed using a PADAC assay. Three clones were identified
35   that expressed elevated levels of protein relative to the

SUBSTITUTE SHEET (RULE 26)

-44-

nonmutated ribosome binding site sequence. These clones
were selected and the nucleic acid sequence of the mutated
ribosome binding site was determined. Details of this
experimental protocol are provided in Example 3.

5   In yet another preferred embodiment of this
invention, a method is provided for creating and
identifying novel regulatory regions in eukaryotes. Example
4 details a method for producing signal peptide libraries
in mammalian cells for improved protein expression. Known
10   eukaryotic signal sequences are disclosed and reviewed by
Von Heigne, G., *Eur. J. Biochem.* 133:17-21, 1983. Figure 7
illustrates an exemplary vector contemplated for use in
this invention. The vector uses the Immunoglobulin heavy
chain promoter to initiate transcription of the desired
15   gene inserted downstream from the promoter. The gene to be
expressed along with the regulatory region to be
mutagenized is inserted into the vector. The regulatory
region is deleted and a random or biased regulatory
sequence is inserted into the vector. Example 4 teaches the
20   mutagenesis of the signal sequence. Following the
incorporation of the random signal sequence into the
eukaryotic vector, the vector is introduced into suitable
eukaryotic cells that are capable of expressing protein
from the expression vector. Example 4 employs the vector
25   of Figure 7 in SP 2/0 cells. Vectors can be introduced
into cells using those methods known in the art. Example 4
employs electroporation to mediate transfection. However,
those with skill in the art will be readily able to select
an appropriate method for introducing the vector into the
30   eukaryotic cells. Positive cells are preferably selected
by growing the cells in neomycin. In Example 4, the
relative level of protein expression is assayed using a
Western dot blot. Positive transfected cells can be
further quantitated for protein expression by ELISA as
35   compared with cells expressing protein under the control of

SUBSTITUTE SHEET (RULE 26)

-45-

the native regulatory sequence. Cells expressing equal or
greater levels of protein expression as compared with cells
containing the native regulatory sequence are subjected to
DNA sequencing.

5   In another preferred embodiment of this
invention, the methods of the invention are used to
generate targeting sequences to target proteins to
particular cellular locations. The terms "protein
targeting sequence" and "protein trafficking sequence" are
10   used interchangeably in this disclosure. Both prokaryotic
and eukaryotic systems studied to date indicate that
sequences associated with protein targeting can occur in a
variety of locations within a protein sequence. For
example, in bacteria, proteins with N-terminal signal
15   sequences are secreted into the periplasmic space or are
inserted into the outer membrane using a secretion
apparatus defined by a series of sec genes (Bassford, et
al., *Cell* 65:367, 1991). Other proteins lacking signal
sequences are also secreted into the extracellular space in
20   *E. coli* and some of these proteins contain secretion
signals located at the C-terminus of the protein
(Delepelaire, P. et al, *J. Biol. Chem.* 265:17118, 1990).
These secretory proteins lack a classical hydrophobic
signal peptide. Instead, the proteins contain a consensus
25   region of 200 amino acids containing highly conserved
stretches of amino acids that are postulated to confer
secretory activity.

             Similarly, targeting sequences are located within
a variety of locations in eukaryotic proteins. For
30   example, interleukin proteins lack hydrophobic leader
peptides and their targeting sequences are believed to be
located in the C-terminal portion of the protein precursor
(Marck, et al., *Nature*, 315: 641, 1985). *E. coli*
expression of Interleukin-1 results in the translocation of
35   the protein to the periplasmic space, while expression in

SUBSTITUTE SHEET (RULE 26)

-46-

yeast results in expression of the protein in the cytosol. Particular protein sequences also are implicated in mitochondrial targeting (see Zara, et al., *J. Biol. Chem.* 267:12077-12081; Adrian, et al., *Mol. Cell. Biology* 6:626-

5   634, 1986 and Horwich, et al, *Cell* 44:451-459). Similarly in both prokaryotes and eukaryotes, post-translational modifications, such as glycosylations (eukaryotes), the addition of lipids, or the like, have been shown to be important in protein translocation and targeting within a

10 cell. It is generally believed that specific regions within the protein sequence are important determinants for post-translational modifications within a cell. Kuchler, et al, *Endocrine Reviews* 13:499-514, 1992 observed that the C-terminal modification of a chimeric IL-1_::bacterial a-

15 factor protein resulted in exclusive plasma membrane association. Others have demonstrated that a combination of post-translational modifications of the C-terminal region of a heterologous protein with the presence of a polybasic domain, positioned near the C terminus, result in

20 the targeting of heterologous protein to the plasma membrane in MDCK cells (Hancock, et al. *EMBO J.* 10:4033, 1991). In another series of experiments, Bakke, et al. (*Cell* 63:707-716, 1990) identified the region of a protein and the particular sequence responsible for sorting a

25 plasma membrane protein from the endosomal vesicles to the plasma membrane. For a review of the location and sequences of known eukaryotic sorting sequences see Schwartz, *Ann. Rev. Immunol.* 8:195-229, 1990.

The ability to target protein to a particular

30 cell location is an important goal in gene therapy strategies. Advantageously, the method described in the claims and below and detailed in Example 5, permits one with skill in the art to introduce a particular nucleic acid sequence encoding protein into a cell and to select

35 cells expressing protein in the desired cell location. A

SUBSTITUTE SHEET (RULE 26)

-47-

comparison of the native protein sequence with the mutated
protein sequence that targets the protein to the desired
cell location permits the identification of the novel
sequence that directs protein targeting.

5   As a first step for practicing the trafficking or
targeting method of this invention, a particular region of
nucleic acid containing a protein trafficking/targeting
signal or a postulated protein trafficking/targeting signal
is identified. As noted, trafficking signals can be
10   positioned in a variety of locations within a protein
sequence. They have been identified with the leader
sequence as well as within the C-terminal portion of a
protein. Those with skill in the art of molecular biology
will be able to search the literature for recognized
15   regions of eukaryotic or prokaryotic protein sequences that
are postulated to affect protein trafficking and/or
targeting. Equivalent regions within the nucleic acid
sequence are then identified.

          Next, a nucleic acid vector is prepared that
20   contains the nucleic acid sequence encoding the protein
trafficking signal to be mutagenized and a nucleic acid
sequence encoding at least one protein. The nucleic acid
vector may be suitable for directing eukaryotic or
prokaryotic expression of the protein, and those with skill
25   in the art of molecular biology will be readily able to
select and construct an expression vector suitable for this
invention. Once the vector is prepared, the nucleic acid
sequence containing the protein trafficking signal is
deleted from the vector.

30   As a next step for practicing the methods of this
invention, a pool of random oligonucleotides are produced
that are suitable for use in a polymerase chain reaction.
These oligonucleotides are incorporated into the position
in the vector previously occupied by the protein
35   trafficking signal to generate a pool of mutagenized

SUBSTITUTE SHEET (RULE 26)

-48-

vectors. Suggested methods for the steps of this invention
are discussed _supra_ as they relate to methods for creating
nucleic acid regulatory sequences and specific examples of
the suggested methods are provided in the Examples below.

5   The mutagenized vectors are introduced into a
cell sample using electroporation, membrane
permeabilization, CaCl$_2$ precipitation, viral vectors, or the
like. Those with skill in the art will use their preferred
methods for introducing nucleic acid vectors into a cell
10   sample.

As a next step for practicing the methods of this
invention, the cells are assayed to determine the location
of the protein encoded by the nucleic acid vector in the
cell sample. Methods to determine the localization of a
15   protein within a cell are known in the art. For example,
if antibodies are available that react with the protein
encoded by the nucleic acid vector, then these antibodies
can be labelled with a fluorescent marker, colloidal gold,
or the like. Cell samples expressing the protein can be
20   membrane permeabilized for fluorescent antibody analysis or
cells can be fixed and sectioned for protein localization
studies using electron microscopy in the presence of
colloidal gold. In addition, cell samples can be
fractionated and individual organelles separated and
25   individually tested for the presence of the protein. Cells
expressing the protein in the desired cell location are
identified by an assay and the nucleic acid sequence of the
novel protein trafficking/targeting signal is determined
and compared to the native sequence.

30   Particular embodiments of the invention will be
discussed in detail in the following examples and reference
will be made to possible variations within the scope of the
invention. There are a variety of alternative techniques
and procedures available to those of skill in the art which

SUBSTITUTE SHEET (RULE 26)

-49-

would similarly permit one to successfully perform the
intended invention.

Example 1

5   Creation of Novel Signal Sequences in *E. Coli*
            This example illustrates a method for producing
novel signal sequences for optimized expression of
eukaryotic or prokaryotic protein in prokaryotic cells.
            The sCCHA225 β-lactamase fusion protein was

10 cloned as follows: sCCHA225 was amplified from the
construct pUCHAsc2 using a 5'-terminal primer (B444,
designated SEQ ID NO: 1, see Figure 1) containing an Xba I
restriction site and a 3' primer (B456, designated SEQ ID
NO: 2) containing DNA coding for the (gly$_4$ser)$_2$ sequence and

15 a Class IIS restriction endonuclease restriction site, Bsa
I. The designation "2" in the name pUCHAsc2, indicates
that the ampR gene was a mutation eliminating the BSA1
site, which was accomplished by a single round of EIPCR
using primers 939 (SEQ ID NO: 3) and 940 (SEQ ID NO: 4) as

20 described in Example 2). β-lactamase was amplified from
pUCHAsc2, which contains a modified copy of β-lactamase
lacking the Bsa I restriction site, but containing the
native amino acid sequence. A 5'-terminal primer (B457,
designated as SEQ ID NO: 5) containing a Bsa I restriction

25 site and nucleotides coding for (gly$_4$ser)$_2$ were used. The
3'-overlap primer (B447, designated as SEQ ID NO: 6)
introduced 3 consecutive stop codons and a Bam HI
restriction site at the 3' end of the β-lactamase gene. The
use of overlapping primers, B456 and B457, created a 10

30 amino acid (gly$_4$ser)$_2$ sequence between the single chain
antibody and β-lactamase domains (Figure 1). The resulting
PCR products were processed with Bsa I, ligated and the
ligation products were used as template for a second PCR
reaction using the external primers B444 and B447. Ligation

35 conditions are provided below. The 1.6 kb PCR product was

SUBSTITUTE SHEET (RULE 26)

-50-

then gel purified and subcloned into the expression plasmid
pCCHAscl cut with Xba I and Bam HI. pCCHAscl is a pUC based
plasmid that constitutively expresses a chloramphenicol
resistance gene and contains an expression cassette having
5   a Lac promotor driving expression of the CHA 255 single
chain antibody through the use of an Omp A derived ribosome
binding site and signal sequence. The Omp A sequence was
created from primers using the sequence provided in the
literature (Movva, et al. *J. Mol. Biol.* 147:317-328, 1980).
10   The resulting plasmid pCCHAscl-ampRV2 (see Figure 1) was
verified by restriction digestion and DNA sequencing.
Double stranded dideoxy sequencing was performed on a
Dupont Genesis 2000, using the DuPont Genesis 2000
sequencing kit according to the manufacturer's
15   instructions. Post gel processing was done with the Base
Caller 5.0 program (DuPont, Boston, MA).
        Oligonucleotides used to produce the fusion
protein construct are shown in Figures 1 and 2.
Oligonucleotides were synthesized in an Eppendorf Synostat
20   D automated DNA synthesizer (Madison, Wisconsin).
Oligonucleotides used for library EIPCR were synthesized
with the 5'-trityl group on and were purified with a
Nensorb Prep column (New England Nuclear, Tozer, MA)
according to the manufacturer's instructions.
25   Oligonucleotides containing random nucleotide incorporation
were prepared by selecting a mixed base option according to
the software provided by the manufacturer.
        Standard ligations contained 0.1-1.0 μg DNA, 1x
ligation buffer and 1-2 μl (400,000 units) of T4 DNA ligase
30   (New England Biolabs, Tozer, MA) in a 20 μl volume. The
ligation reactions were incubated at room temperature for
one hour. Mass estimates for chemically synthesized
oligonucleotides were obtained spectrophotometrically. In
all other cases, DNA mass estimates were made by visual
35   comparison of samples run on agarose gels with a 1 kb

SUBSTITUTE SHEET (RULE 26)

-51-

ladder DNA standard (GIBCO/BRL, Gaithersburg, MD) loaded to approximate 100 ng DNA per band.

The construct pCCHAsc1-ampRv-2 was introduced into *E. coli* DH-10B (GIBCO-BRL, Gaithersburg, MD). Bacteria

5   were propagated at 30°C in Terrific Broth medium (Sambrook, et al. 1989. supra) or on agar plates containing 34 μg/ml chloramphenicol and 100 μg/ml ampicillin. Positive clones were selected from the ampicillin plates for further study. Assays to monitor the level of protein expression in the

10 bacteria are described below. The completed vector provides the template for EIPCR-based construction of the signal peptide library.

For EIPCR, the PCR reactions contained 0.2 ng the template, 0.5 _M of each primer, 1X Taq buffer (Perkin-

15 Elmer Cetus, Norwalk, CT), 200 _M of each dNTP, 1.61 mM MgCl$_2$, and 5 units of Taq DNA Polymerase (AmpliTaq™, Perkin-Elmer Cetus) in a total volume of 100 μl. The addition of MgCl$_2$ beyond 1.61 mM up to 3.11 mM is often used as a variable that when modified impacts the yield of DNA.

20 Templates were amplified in a Perkin Elmer Thermocycler (Norwalk, CT) using a "hot start" according to manufacturer's instructions (3 min. at 94°C) followed by 30 cycles of 94°C (1 min.) denaturation, 50°C (1 min.) annealing and 72°C (2 min.) extension. A 5 second auto

25 extension was used. Extension was completed with an additional 10 minute incubation at 72°C.

The EIPCR template DNA is similar in size to the desired EIPCR product, hence they could copurify by agarose gel electrophoresis during further manipulations. While

30 the circular template is present at a much lower concentration than the EIPCR product, it will electroporate at high efficiency in bacteria thereby resulting in the contamination of the library with wild-type clones. The data presented in Table 1, below, demonstrate the problem

35 of wild type clone contamination in library mutagenesis.

SUBSTITUTE SHEET (RULE 26)

-52-

When pCCHAscl-ampR was used as a template, only the wild-
type sequence was recovered from the library under
selective conditions.

5   To render the expression construct nonfunctional
and eliminate the appearance of functional template in the
EIPCR product, the entire Omp A signal sequence was deleted
from pCCHAscl-ampRV-2 in a single EIPCR reaction (see
Figure 2A). Two primers, B593 (SEQ ID NO: 7) and B594 (SEQ
ID NO: 8), were used in an enzymatic inverse polymerase
10   chain reaction to remove the signal sequence. The
oligonucleotides hybridized to opposite strands of the
plasmid and in a PCR reaction yielded a linear DNA that is
slightly smaller than the original template. This linear
DNA contains at its termini a 14 bp extension containing
15   the recognition site for the class IIs restriction site Bsa
I. The terminal extensions of DNA were removed by
digestion with Bsa I, yielding compatible cohesive ends
that efficiently undergo an intramolecular ligation. The
resulting plasmid, pCCHAscl-ampR_SP, is a 3.2 kB pUC-
20   derived plasmid which contains a single chain CHA225-
β-lactamase fusion protein absent the Omp A signal peptide,
and this vector confers resistance to chloramphenicol.
Removal of the signal sequence resulted in ampicillin
sensitivity. Thus, any pCCHASC1-ampR_SP clones that
25   appeared in the final EIPCR library, were eliminated during
selective expression to yield the desired mutagenized
clones (see Table 1, below).

The random oligonucleotides, created as described
above, were incorporated into pCCHAscl-ampRV2_SP using a
30   polymerase chain reaction. Figure 2b illustrates the use
of EIPCR to create a random signal peptide library.
Divergent oligonucleotide primers were used to amplify the
DNA and to incorporate mutant sequences. The forward
primer, B524 (SEQ ID NO: 9), annealed to the DNA
35   corresponding to the N-terminal region of the scCHA

SUBSTITUTE SHEET (RULE 26)

-53-

antibody. The reverse primer, B525 (SEQ ID NO: 10),
incorporated the initiator codon and 21 random codon
triplets. The reverse primer's annealing location is
illustrated in Figure 2b. Both the forward and reverse
5 primers contained the Bsa I restriction site near their
termini. The Bsa I restriction site is present at both
ends of the linear EIPCR product and was removed by
digestion, yielding compatible cohesive ends. The linear
PCR product was then ligated and electroporated into *E.*
10 *coli.*

The infrequent occurrence of an active signal
peptide in a library of random signal peptides necessitated
several considerations for library construction. First, to
ensure a large enough library size and complexity we used a
15 higher than normal amount of template DNA in our EIPCR
reactions. Typically EIPCR reactions contain only 0.5 ng
of template DNA per 100 μl reaction (Stemmer, et al.
supra). Increasing the amount of template to 25-50 ng per
100 μl reaction resulted in more EIPCR product and a larger
20 library size. As an example, when pCCHAscl-ampRV2 was used
as a template, 5 ng of plasmid yielded only 0.25 μg of PCR
product. When 50 ng of template was used, 5 μg of PCR
product was obtained. It is contemplated that the activity
of different primer-template combinations will be optimized
25 within a range of template concentrations from about 0.1 to
50 ng of template per 100 μl reaction. Those with skill in
the art of molecular biology can readily determine the
optimal template concentrations as well as adjust the
concentration of $MgCl_2$ to maximize PCR reaction product.
30 Advantageously, the deletion of the critical regulatory
sequence from the template prior to the incorporation of
random sequence enables one to use greater amounts of
template than those typically used in standard PCR
reactions without concern for amplifying nonmutated
35 template. EIPCR reaction conditions for the incorporation

SUBSTITUTE SHEET (RULE 26)

-54-

of random oligonucleotides were the same as for the cloning reactions.

Table 1

Library Template	Template DNA/100 μl	CFU/μg DNA		% Wildtype (n)
		CAM	CAM/AMP	CAM/AMP
pccHAscl-ampR	20ng	20.4×10^6	4511	100% (8)*
pccHAscl-AmpRΔSP	50ng	3.9	1904	0 (10)**

CAM (chloramphenicol) = 34 μg/ml
AMP (ampicillin) = 50 μg/ml
n=number of samples evaluated
*=determined by restriction mapping
**=determined by DNA sequencing

The 100 μl PCR mixture was extracted with an equal volume of phenol-chloroform-isoamyl alcohol (24:23:1) and was then precipitated using 2 volumes of ethanol. The protruding termini of the PCR end products were filled in and digested according to the method of Stemmer (*Biotechniques*, 1992 supra). Briefly, the pellet was resuspended in 70 μl H_2O and 10 μl of 10X Klenow buffer (NEB, Tozer, MA), 10 μl dNTP mix (2.5 mM each dNTP), 5 μl of DNA Polymerase I (large fragment: Klenow), and 5 μl of T4 DNA polymerase were added. The reaction was incubated at 37°C for one hour followed by phenol extraction and ethanol precipitation as disclosed above. Blunt-ended EIPCR fragments were then exhaustively digested in a 100 μl reaction containing 50 units of Bsa I at 60°C for four hours. The Bsa I digested fragment was purified away from both the PCR primers and the short terminal DNA fragment by

SUBSTITUTE SHEET (RULE 26)

-55-

centrifugation through a Biospin 6 column (BioRad, Richmond, CA) according to the manufacturer's protocol. The purified Bsa I digested fragment was then phenol extracted and ethanol precipitated as described earlier and

5   ligated under standard conditions with the exception that the fragments were ligated at higher concentrations of T4 DNA ligase (106 units/μl) and incubation was at 12°C overnight.

         Ligated or control DNA (pUC 19, GIBCO/BRL) was

10 precipitated in 2 volumes of ethanol and resuspended in 20 μl of TE (10mM Tris, pH 7.4, 0.1 mM EDTA, pH 8.0). 10 μg of yeast tRNA (GIBCO/BRL) was added as a carrier and did not affect electroporation efficiency. The DNA was electroporated in 1-5 μl aliquots into electrocompetent

15 DH10-B MAX cells (BRL, Bethesda, MD) according to the manufacturer's instructions. pUC19 DNA (0.01 ng) was electroporated to monitor the electroporation efficiency.

         Under conditions to select for the plasmid marker (34 μg/ml chloramphenicol), a library of 3.9 x 10⁶ colony

20 forming units was obtained from approximately 1 μg of ligated DNA (Table 1). Under dual selective conditions (34 μg/ml chloramphenicol + 50 μg/ml ampicillin) an equivalent amount of DNA yielded 1904 cfu, which represents a positive rate of 0.05%. Clones showed a variety of colony sizes

25 ranging from 0.2-2 mm.

         When greater concentrations of ampicillin were used, the fraction of positive clones decreased (see Table 2, B). In contrast, the plating efficiency of the control plasmid using CAM selection pCCHAscl-ampR (A), which

30 contains the wild Omp A signal, did not decrease appreciably at higher concentrations of ampicillin. The libraries formed under higher concentrations of ampicillin presumably contain clones with leader peptides capable of higher levels of secretion. In these experiments 10 μg of

35 pCCHAscl-ampR was electroporated into 20 μl

SUBSTITUTE SHEET (RULE 26)

-56-

electrocompetent *E. coli* DH10B and grown for one hour at
37°C in 1 mL S.O.C. (Sambrook, et al. supra). The cells
were centrifuged briefly and resuspended in 400_1 S.O.C. A
10^{-4} dilution was prepared in S.O.C. and 0.1 ml aliquots
5   were plated on TB plates containing 34 μg/ml
chloramphenicol with the indicated concentration of
ampicillin. This transformation strategy was compared to a
transformation using 4 μl (approximately 1 μg) of signal
peptide library DNA electroporated into 80 μl
10   electrocompetent *E. coli* DH10B and processed as described
for the control pCCHAscl-ampR plasmid except that the 10^{-4}
dilution step was eliminated.

Table 2

15

Restriction of the signal peptide library on increasing
concentrations of ampicillin

	Clone/Library	[amp] (μg/mL)	cfu/plate	(+) cfu by ^{111}In screen	% positive
A	pCCHAsclampR	50	254	199	78.3
		100	218	149	68.3
		200	319	201	63.0
		400	211	74	35.1
B	SP Library	50	640	107	16.7
		100	313	98	31.3
		200	118	42	35.5
		400	12	6	50.0

The library size was determined by
electroporating 1 μl of library DNA and plating 90% of the
electroporated cells onto media containing 34 μg/ml
35   chloramphenicol and 100 μg/ml ampicillin and by plating
serial dilutions of the remaining 10% of the electroporated
cells onto media containing only chloramphenicol (34

-57-

μg/ml). Under the latter conditions all clones produced
colonies. The quality of the resulting signal peptide
library was verified by DNA sequencing of the mutated
region of four nonselected clones. In all clones the

5   mutations were correctly incorporated. The composition of
the mixed bases was 31% A, 19% G, 25% C and 25% T where the
sample number was 119.

Library clones that were able to grow on 50 μg/ml
or greater ampicillin were further verified by colony lift

10 and by hapten binding assays. One possible complication of
ampicillin selection can be deletion of all or some of the
scCHA insert. Another possibility is that the EIPCR
reaction could incorporate mutations in the antibody
sequence and thereby eliminate antibody activity. In fact

15 some errors were identified when individual clones were
later sequenced. To examine these possible complications,
two approaches were used. First, digestion of DNA from
positive clones (grown with 200 μg/ml ampicillin) with Xba
I and Pst I indicated that about 70% of the clones

20 contained the full length 800 scCHA insert.

Second, a large number of colonies were grown
under selective conditions, transferred to nitrocellulose
and probed with the radioactive hapten ^{111}In-EOTUBE. Labeled
chelate was prepared and colony lifts performed as

25 described by Stemmer, et al. (*BioTechniques* 14:256-265,
1993). Briefly, colony lifts of 23cmx23cm plates with 0.3-
1 x 10[5] colonies were prepared using BA83 nitrocellulose
filters (Schleicher and Schuell, Keene, New Hampshire).
The filters were blocked by incubation in 3% non-fat milk

30 in 25 mM Tris-HCl pH 7.5 for 10 minutes, washed with 25 mM
Tris, followed by incubation in 25 mM Tris containing 50
°Ci of chelated [111]Indium per filter for 1 hour at room
temperature. The filters were then washed with 25 mM Tris
for a total of 15 minutes, dried and exposed to Kodak X-

35 Omat AR autoradiography film for several hours.

SUBSTITUTE SHEET (RULE 26)

-58-

Approximately 30-50% of the colonies yielded a
positive signal in this assay, indicating that a
significant number of clones were not expressing active
scCHA antibody (see Table 2, supra). Loss of scCHA
5   activity also occurred in the control plasmid, though at a
lower frequency. Clones yielded positive scCHA signals of
varying intensities. Twelve clones yielding a high
intensity signal were selected for further work and were
purified by a secondary and tertiary colony screen. These
10 positive clones were removed from the primary screen plates
and diluted into fresh media, replated and reassayed. A
single positive colony from this secondary screen was
replated and reassayed in a tertiary screen, which yielded
clonal material, since 100% of the colonies yielded a
15 positive signal. The results of representative colony
lifts shown demonstrated that the scCHA signal was stable
during successive rounds of growth (data not shown).

In order to study protein expression in the
cells, the cells were grown at 30°C in Terrific Broth (see
20 supra) containing 34 μg/ml chloramphenicol and 50 μg/ml
ampicillin in baffled shaker flasks at 250 rpm for 24
hours. Following cell growth, the periplasmic fraction of
E. coli was prepared and isolated using the methods
described by Witholt, et al. (Anal. Biochem. 74: 160-170,
25 1976, hereby incorporated by reference).

Determination of antibody expression levels was
performed as previously described (Stemmer, et al.,
BioTechniques, 1993 supra). The quantitation of the active
fusion protein was based on a comparison to the CHA 255
30 standard. Isolated clones were further characterized to
demonstrate the production of an active fusion protein. To
quantitate the amount of fusion protein recognizing the
labelled hapten, 50 μl of periplasmic extract was incubated
with 50 μl of carrier free ^{111}In-EOTUBE for 30 minutes. The
35 mixture was then centrifuged through an ultrafree column

SUBSTITUTE SHEET (RULE 26)

-59-

(10,000 MW cutoff) and the eluate was counted. The % bound
was expressed as the (input cpm-eluate cpm)/input cpm. The
number of sites (in pmol) was determined from a standard
curve utilizing CHA255 antibody. The data of Table 3
5   indicates that periplasmic supernatants from several clones
bound to the radioactive hapten ^{111}In.

<u>Table 3</u>

10   Hapten binding and β-lactamase activity of protein
                  containing mutant signal peptides

clone	pmol site/ml	ng/ml	ΔAU/min/ml	ng protein/ml
CHA	12	960.0	NA	NA
β-lac STD	NA	NA	3.038	680
ompA Fusion protein	1.13	59.9	1.075	481
E3	0.009	0.45	0.104	45.6
E5	0.076	4.04	0.079	34.66
E6	0.638	33.8	1.409	618.1
E9	0.098	5.17	0.089	39.05
E10	0.009	0.5	0.371	162.76
E12	0.131	6.93	0.697	305.8
E13	0.33	17.47	0.184	80.72
E14	0.36	19.09	0.086	37.7
E15	0.079	4.19	0.074	32.46
E16	0.007	0.4	0.225	98.7
E17	0.012	0.65	0.087	38.16
E18	0.043	2.29	0.083	36.41

SUBSTITUTE SHEET (RULE 26)

-60-

E19	0.032	1.7	0.081	35.53
E20	0.014	0.76	0.074	32.5
E21	0.03	1.58	0.104	45.6
E22	0.035	1.85	0.100	43.9

5

Since library clones were isolated from
ampicillin containing plates, one would predict that
periplasmic supernatants should also contain significant
β-lactamase enzymatic activity. The activity of the
10   β-lactamase portion of the fusion protein was determined by
the quantitative PADAC (2-[(N,N-dimethylanilin-4-
yl)azo]pyridinium 3'-cephalosporin) assay (Meyer, et al.,
Bioconjugate Chem. 3:42-48, 1992, which is hereby
incorporated by reference). 100 μl aliquots of periplasmic
15   supernatant were incubated at 37°C with A_{570} = 0.5 of PADAC
in a 2 ml stirred cuvette and the decrease in the A_{570} was
monitored over time. β-lactamase activities were expressed
as a change in absorbance unit/min/ml and mass
determinations were made based on a commercially available
20   E. coli β-lactamase (Sigma St. Louis, MO) standard of known
concentration assayed under identical conditions. Results
are provided in Table 3. The data reveals that all of the
clones tested had significant β-lactamase activity.
Interestingly, mutant E6 had the highest activity, above
25   that of the control fusion protein. The use of a
β-lactamase standard permitted calculation of the
approximate mass of fusion protein produced.

Based on the data in Table 3, there seems to be
no clear direct correlation of antibody binding activity
30   with β-lactamase activity. For example, some clones like
E13 and E14 showed moderate levels of antibody activity but
very low levels of beta lactamase activity. Also, clones
like E10, E12 and E16, showed moderate levels of beta
lactamase activity while exhibiting very little antibody
35   activity. Only clone E6 showed levels of antibody and

SUBSTITUTE SHEET (RULE 26)

-61-

enzyme activity that were close to that for the control Omp
A clone. Because there is the possibility of obtaining
mutations in any PCR technology that may affect the
activity of antibody or enzymes, we took several of the
5   periplasmic preparations and evaluated them for their
expression levels by Western blotting (Sambrook, et al.
supra) since this method should be less sensitive to
mutation (see Stemmer, et al. BioTechniques, 1993 supra).
Samples were developed with purified rabbit polyclonal
10   antibodies to CHAsFv. The antibodies were obtained from
rabbit polyclonal antiserum. Rabbits were immunized by a
traditional procedure (J. Immunol. 116:1306-1312, 1975)
that employed multiple intramuscular injections of CHA255
monoclonal antibody (100 µg amounts) in Freund's adjuvant.
15   Serum obtained from the rabbits were processed to isolate a
purified polyclonal antibody preparation to CHA255 Fv by
absorption and elution from an immunoadsorbant column that
was conjugated with CHA255 chimeric antibody. The CHA255
chimeric antibody was cloned and produced as described
20   previously (J. Immunol. 145:1200-1204, 1990 and Protein
Expr. and Purif. 2:75-82, 1990). CHA255 chimeric antibody
was conjugated to CNBr preactivated Sepharose 4B according
to the manufacturers instructions (BioRad Labs, Richmond,
CA). The antibody in the rabbit polyclonal antiserum were
25   applied to the column and eluted as described previously
(J. Immunol. 116:1306-1312, 1975). This purified
polyclonal antibody preparation was used in Western blots
in 100 µg/ml concentrations. The results of the Western
Blot experiments indicated that E5 was now the best
30   producer of the mutants tested from Table 3. In this
experiment, the Omp A control clone produced higher levels
of protein than the mutants. Expression of E5 was higher
than that of E6, with E12 and E10 slightly lower in
expression than E6 (data not shown).

SUBSTITUTE SHEET (RULE 26)

-62-

In addition, several of the selected library E
clones were subjected to DNA sequencing to determine the
sequence of the DNA that coded for the signal peptide (see
Table 4). A number of the clones contained stop codons
5 within their sequence. In particular, clone E6, which
directed increased protein expression in the antibody and
enzyme activity tests and was second only to E5 in the
Western blot, contained several stop codons in the signal
sequence.
10

Table 4

AMINO ACID SEQUENCES OF SIGNAL PEPTIDE LIBRARY E

15	LEADER	SEQUENCE		
	OmpA(SEQ ID NO: 11)	MKKTAIAIAV	ALAGFATVAQ	A
	E3 (SEQ ID NO: 12)	MLYPGEMLCM	QNFITLIIEI	YI
	E5 (SEQ ID NO: 13)	MGMYKLRLLI	DCLMLYVCLT	DT
	E6 (SEQ ID NO: 14)	MNILLEN*M*	RTHDMLIDGT	I
20	E9 (SEQ ID NO: 15)	MGMYKLRLL	DCLMLYVCST	DT
	E10(SEQ ID NO: 16)	MTSVCKITVL	ILFXTGTYNI	LG
	E12(SEQ ID NO: 17)	MRNTYV'YXKY	GRTEWLVEQI	NL
	E13(SEQ ID NO: 18)	MvQCERXIYH	YvQRIREYFX	FE
	E14(SEQ ID NO: 19)	MvGICTGIGV	GILLLLLVVG	QV
25	E15(SEQ ID NO: 20)	MLGAMEFKKL	SYAILVLSHT	PT
	E16(SEQ ID NO: 21)	MDQGIKYRLE	KWRGRISLIR	VF
	E17(SEQ ID NO: 22)	MQRRGIRWAY	VIYLSMFFGA	C
	E18(SEQ ID NO: 23)	MLGVLTMPCD	TMVEDLACDY	MS

* = stop codon
30 X = unknown amino acid

Because the EIPCR method involves amplification
of the entire plasmid, there is always the possibility that
one can have mutational errors occurring in any location in
35 the vector of the selected library clones. Thus, it is

SUBSTITUTE SHEET (RULE 26)

-63-

always recommended that after the DNA sequence is
determined for the mutated library segment, the selected
new sequence is recloned using new DNA oligonucleotides
with standard cloning procedures to confirm that the
5   sequence is in fact directing the expected effect
identified from the originally selected mutant sequence. In
the case of the signal sequence library E, we chose to
reclone E5 and E10 signal sequences in place of the natural
β-lactamase signal sequence and compare expression with Omp
10   A driven β-lactamase. To accomplish this, oligonucleotides
containing the new sequences were cloned into the vector
pCLA3ampR (Figure 3). This PUC derived vector contains a
constitutively expressed chloramphenicol gene, LacI, and
the β-lactamase gene. Expression of β-lactamase is driven
15   by the Lac promoter and is controlled by the Omp A RBS
sequence and Omp A signal sequence. The activity of the
Lac promoter is negatively regulated by the LacI gene
product, the Lac repressor.

        Construction of pCLA3ampR E5 was accomplished by
20   PCR using primers B947 (SEQ ID NO: 24) and 940 (5' and 3'
primers respectively, see Figure 9) with the pCLA3ampR
template using standard PCR conditions as disclosed supra.
The product was cut with Xba I and Pst I. The PCR product
was then ligated back into pCLA3ampR, which had previously
25   been digested with Xba I and Pst I. The construction of
pCLA3ampR E10 was accomplished in the same way, except that
the 5' primer was B946 (SEQ ID NO: 25).

        In order to determine the expression level of
β-lactamase, clones were grown in Terrific Broth containing
30   30 μg/ml chloramphenicol in baffled shaker flasks rotating
at 250 rpm for 24 hours at 30°C. IPTG (Boehringer
Mannheim, Indianapolis, IN) was added (final culture
concentration 0.4 mM) to the cultures when the OD_{600} reached
0.8 to induce the expression of the β-lactamase gene. The
35   periplasmic fraction was obtained as described by Witholt,

SUBSTITUTE SHEET (RULE 26)

-64-

et al. (supra). The periplasmic samples were analyzed for
β-lactamase activity using the PADAC assay described
earlier. The results indicated that the E5 and E10 signal
sequences were equivalent to Omp A in their ability to
5 express β-lactamase from the pCLA3 vector (Table 5). It
should be noted that the DNA sequence coding for the signal
peptides of clones E5 and E10 depicted in Table 5, were
determined. The sequence of E10 was as previously
determined, but E5 showed a T to C base change at position
10 8, resulting in a change of the third amino acid of E5 from
a methionine to a threonine. The results in Table 5
indicate that this single amino acid change in E5 did not
adversely affect its ability to function as a signal
sequence.

15

-65-

Table 5

Clone	Exp. 1	Exp. 2	Exp. 3	Average/ (stdev)	% of Control
		AU/min /ml x 10^{-3}			
pCLA3 ampR Control	19	23	22	21.3/ (2.1)	100
RBS 1 short			40		188
RBS 2 short	46	48	35	43/(7.0)	202
RBS 2 long	40	47	38	41.7/ (4.7)	196
RBS 8 short	57	71	59	62.3/ (7.6)	293
RBS 12	39	47	36	41.7/ (3.5)	196
Signal peptide E 5		20	22	21/(1.0)	99
Signal peptide E 10		21	23	22/(1.0)	103
Blank		0	0	0	0

The library generated E5 and E10 signal sequences were then compared for structural properties with selected

SUBSTITUTE SHEET (RULE 26)

-66-

signal sequences derived from E. coli proteins (Omp A,
DGAL, LamB, ELT6, Omp T, Pel B, PhoA). These signal
sequences, as well as a variety of others are available
from the GenBank database. The program Protean (DNAstar,
5   Madison, WI) was used to evaluate alpha helix vs beta sheet
regions (Garnier-Robson and Chou-Fasman methods) as well as
the hydrophilicity index (Kyte-Doolittle method). Both E5
and E10 displayed the characteristic hydrophobic core of
amino acids seen in all of the *E. coli* derived signal
10   sequences (Figure 4). Interestingly, the E10 sequence
contained no alpha helix regions. Alpha helix content
varied with all of the native *E. coli* signal sequences
evaluated. Also, an interesting feature of the E5 signal
sequence was that the final 3 residues of the sequence,
15   Thr, Asp, Thr, correspond to those designated in the art as
a -3,-1 box which provides the recognition sequence for
cleavage by the *E. coli* signal peptidase (Protein
Targeting, supra). The sequence in the -3,-1 box for E5
differs from all known *E. coli* sequences. The presence of
20   this unique -3,-1 sequence is not detrimental to the
ability of E5 to export a functional β-lactamase protein to
the periplasm.

To demonstrate that the fusion proteins expressed
with mutant signal peptides were processed to the correct
25   size, periplasmic extracts were fractionated on SDS-
polyacrylamide gels. The proteins were transferred to
nitrocellulose and probed with an anti-CHA Fv antibody.
Results from the Western blots indicated that the fusion
proteins isolated from the bacterial clones were _53 kD.
30   This is close to the predicted size of the fusion protein.

Example 2
Production of Novel Signal Sequences of Variable Length
Figure 5 illustrates the use of primers B755,
35   B756, B757, and B758 (corresponding to SEQ ID NOS: 26 - 30

SUBSTITUTE SHEET (RULE 26)

-67-

respectively) that were used in EIPCR reactions to produce
signal peptide libraries of varying length. These
libraries were constructed using primer B755 and one of the
following primers: B756, corresponding to the random 12 aa

5   signal peptide library; B757, corresponding to the random
16 aa signal peptide library; and B758, corresponding to
the random 20 aa peptide library. In this example, the
varying length libraries were constructed such that the
random signal sequences were incorporated directly onto the

10   amino terminus of the gene sequence encoding β-lactamase
(amp-r). The resulting plasmids are diagrammed in Figure 5
as pCampR SP-12, pCampR SP-16 and pCampR SP-20.

Following EIPCR, and using the methods outlined
in Example 1, several clones from each library were

15   prepared and the DNA was analyzed by restriction
endonuclease digestion to verify the predicted size of the
constructs. In the table below, active clones are defined
as those clones which grow on 34 μg/ml chloramphenicol plus
10 μg/ml ampicillin. The following percentages of active

20   clones were identified. Active clones are defined as those
which grown on 34 μg/ml chloramphenicol plus 10 μg/ml
ampicillin):

Library	% Active Colonies
12 aa	0.0028%
16 aa	0.0057%
20 aa	0.0141%

30   The clones were also analyzed by PADAC assay
(outlined in Example 1) and clones with high β-lactamase
activity were selected for sequencing. The results
indicated that as the length of the signal sequence
increases, the percentage of active clones within the

35   library also increases. Moreover, while it is possible to

SUBSTITUTE SHEET (RULE 26)

-68-

obtain shorter signal sequences that were active, signal
sequences of at least 20 aa have an increased likelihood of
being functional.

5   Example 3
     Creation of Novel Ribosome Binding Sites in *E. coli*
             To apply this invention to a second regulatory
     region, we deleted the DNA sequence in an expression vector
     coding for a ribosome binding site (RBS), and generated a
10   random DNA library in its place. Since the RBS sequence
     initiated translation of β-lactamase, library clones which
     contained a useful RBS sequence were selected by growing
     the library on ampicillin. The RBS library was prepared in
     the pCampR vector (Figure 6). This is a pUC derived vector
15   which has a constitutively expressed chloramphenicol
     resistance gene and expressed β-lactamase under the control
     of the Lac promoter. In addition, translation of the
     β-lactamase gene was accomplished using a ribosome binding
     site (GAGG) from the Omp A gene (Omp A sequence from -18 to
20   -1) followed by the Omp A signal sequence (+1 to +63) to
     direct the β-lactamase protein to the periplasm. The
     pCampR vector has a wild type DNA β-lactamase sequence. In
     order to use the Bsa I enzyme for the EIPCR library
     reaction, a Bsa I site in the β-lactamase gene was first
25   removed. Removal was accomplished by standard EIPCR
     methodology (disclosed above) using the oligonucleotide
     primers 939 (SEQ ID NO: 3) and 940 (SEQ ID NO: 4) (see
     Figure 6). The resulting vector was designated pCampR_BSA.
             Next, the entire region containing the Omp A
30   derived ribosome binding site was removed from pCampR_BSAI.
     Sixteen bases were removed from position -16 to -1 by a
     single round of EIPCR using standard conditions and primers
     B667 (SEQ ID NO: 30) and B668 (SEQ ID NO: 31) (See Figure
     6). The resulting vector was designated pCampR_RBS and was
35   unable to grow on plates containing 100 μg/ml ampicillin.

SUBSTITUTE SHEET (RULE 26)

-69-

To generate the RBS library, pCampRΔRBS was used as the
template in an EIPCR reaction containing primers B765 (SEQ
ID NO: 32) and B766 (SEQ ID NO: 33) (See Figure 6). Primer
B766 contained a sequence of 16 NTPs (25% mixture of each
5   base) so as to exactly replace the prior deleted 16 base
sequence with a random DNA sequence. The EIPCR library
reaction was accomplished using the EIPCR conditions as
disclosed by Stemmer, et al. (*BioTechniques* 13:114-220,
1992). This reaction included low amounts of template
10   (0.5-1 ng/100 μl reaction mixture). This is in contrast to
Example 1, which required much larger amounts of DNA
template (up to 50 ng/100 μl reaction mixture). The amount
of template DNA that needs to be used in a particular EIPCR
library reaction can be experimentally determined by those
15   with skill in the art. As illustrated in Example 1,
deletion of the entire functional region of the DNA prior
to introducing the library sequence eliminates the concern
for wildtype clones contaminating the library and enables
the use of large amounts of template DNA for EIPCR, if
20   indeed large amounts are required.

      The EIPCR reaction material was processed to fill
in the ends with Klenow and T4 DNA polymerase, cut with Bsa
I, and ligated as described in Example 1 except that after
the Bsa I digest, the DNA was electrophoresed on an agarose
25   gel and the band containing the DNA was cut out and
processed using GeneClean (Bio101, San Diego, CA),
according to the manufacturer's instructions, to separate
the digested from undigested fragments.

      The ligated RBS library DNA was electroporated by
30   adding 1 μl (50 ng) of DNA to 20 μl of electrocompetent
DH10-B MAX cells (BRL, Bethesda, MD) according to the
manufacturers instructions. Under conditions to select for
the plasmid marker (34 μg/ml chloramphenicol), the library
size was estimated to be 1.7×10^6 colony forming units/μg
35   ligated DNA. Under dual selective conditions (34 μg/ml

-70-

chloramphenicol + 50 µg/ml ampicillin), an equivalent
amount of DNA was estimated to produce 24,040 colony
forming units (cfu), which indicates that 1.4% of the
colonies contained potential mutants. When greater
5   concentrations of ampicillin were used, the fraction of
positive colonies on the double selected plates decreased
precipitously (Table 6). In contrast, the plating
efficiency of the control plasmid, pCampRΔBSA, did not
decrease appreciably until 400 or 800 ng/ml ampicillin was
10   used. Referring to Table 6, sixteen library colonies from
the dual selective plates (800 µg/ml ampicillin + 34 µg/ml
chloramphenicol) were picked and inoculated into LB media
and grown overnight. In the morning, cells were harvested
and periplasmic fractions were produced according to the
15   above referenced methods. PADAC as says to measure
β-lactamase activity showed a wide range of activity (Table
7), with several clones exhibiting activities higher than
what was obtained with the wildtype Omp A RBS in the
pCampRΔBSA vector. After several repeat cultures and PADAC
20   assays, it was clear that RBS clones 1, 8 and 12 showed
higher levels of expressed β-lactamase activity than the
wildtype Omp A RBS clones pCampRΔBSAI (see Table 7).

-71-

Table 6

ELECTROPORATION COLONIES

[AMP] (μg/ml)	RBS Library	Control pCamp ΔBSA
50	5000	1894
100	760	2007
200	81	1564
400	41	914
800	22	211

Table 7

BETA LACTAMASE ACTIVITIES (PADAC) FOR RIBOSOME BINDING SITE
LIBRARY CLONES

	Selected *		Not Selected **	
Clone#	Activity*** (in ΔAU/min/ml)		Clone#	Activity* (ΔAU/min/ml)
RBS 1	5.58♦		RBS 11	1.95
RBS 2	1.65♦		RBS 12	5.18
RBS 3	2.72		RBS 13	0.012
RBS 4	3.51♦		RBS 14	0.021
RBS 5	3.78		RBS 15	0.016
RBS 6	2.09♦		RBS 16	0.005
RBS 7	0.95			
RBS 8	12.76♦			
RBS 9	2.35♦			
RBS 10	3.54♦			

SUBSTITUTE SHEET (RULE 26)

-72-

*   clones derived from plates containing 800 µg/ml AMP,
      34 µg/ml CAM

**   clones derived from plates containing 34 µg/ml CAM

5   *** PADAC activity for the wildtype ompA RBS sequence
      (pCampRBSAI) is 3.31 _AU/min/ml

♦   These clones had an insertion of 17 bases added 3' of
      the library RBS sequences as measured by an enlarged
      Xba I, Eco RI restriction fragment

10

      Purified DNA from the selected RBS clones were
initially evaluated for key restriction sites. It was
noticed that some of the library clones (RBS 1, 2, 4, 8-10)
had acquired an unexpected mutation during formation of the
15 library. Although it is not unusual to pick up errors
during PCR techniques, it is important to note that the
errors incorporated into the constructs disclosed below had
no impact on the activities of the RBS sequences that were
isolated from the library. This is known with certainty
20 since we later recloned the selected sequences without the
errors and found that the improved RBS sequences were as
active as they were in the initial clones from the library.
      The error we found resulted in an insertion of
about 17 nucleotides arising from incomplete Bsa I
25 digestion of one of the ends of the PCR library reaction,
causing a blunt ended ligation. The insertional error did
not appear to be related to selection conditions of the
library, since it was seen in both the amp/CAM selected
(50%, N=24) and CAM only selected clones (19%, n=16). A
30 comparison of β-lactamase activities of the ampicillin
selected clones also showed no obvious correlations between
the insertion error and the level of beta lactamase
activity produced (Table 7).
      In order to determine if the insertional mutation
35 played a role in the RBS clones exhibiting increased levels

SUBSTITUTE SHEET (RULE 26)

.-73-

of expressed β-lactamase, the DNA coding sequence for the
RBS regions in four clones were determined by DNA
sequencing and new oligonucleotides containing these
sequences were synthesized to enable expression analysis in
5   a new vector that had not undergone EIPCR. The DNA
sequence for RBS clones #1, 2, 8 and 12 are provided in
Figure 8. As noted above, for clones 1, 2 and 8 there was
an additional insertion due to improper cutting of the
terminus of the EIPCR primer (inserted nucleotides shown in
10   bold). In clones 1, 2, and 8, the full length RBS sequence
containing the insertion is herein referred to as RBS long,
while the RBS sequence without the insertion is called RBS
short. DNA oligonucleotide primers coding for resistance,
RBS 2 long and short, RBS 8 short and RBS 12 were
15   synthesized. Their sequences are shown in Figure 9.
         The vector used to retest the activity of these
various new RBS sequences was pCLAampR (Figure 3),
disclosed above. The various clones were produced by
performing a standard round of PCR in which the various
20   primers were included and pCLAampR was used as the template
(Figure 8). The PCR products were filled in with Klenow
and T4 DNA polymerase, followed by restriction digestion
using Xba I and Pst I. The products were cloned into the
pCLAampR vector that had been previously cut with the same
25   restriction enzymes. The DNA was ligated using the methods
provided in Example 1, and DH10B bacteria were transformed
using the electroporation procedures disclosed above.
Clones were selected for growth on chloramphenicol and
grown overnight in Terrific Broth media + 34 μg/ml
30   chloramphenicol. The next day, cultures were seeded at
$OD_{600}=0.1$ in LB media + 34 μg/ml chloramphenicol and grown at
30°C to $OD_{600}=0.8$. IPTG was then added to a final
concentration of 0.4mM. The cultures were grown for an
additional 24 hours at which time the cells were harvested
35   and periplasmic fractions were produced as described in

SUBSTITUTE SHEET (RULE 26)

-74-

Example 1.   The periplasmic samples were next subjected to
PADAC assay for β-lactamase activity determination.   The
results from three separate sets of cultures are provided
in Table 5 and indicate that in the Lac regulated vector
5   all the RBS library sequences shown in Table 5 were more
active than the wild type Omp A RBS sequence.   Also, the
same level of β-lactamase expression was obtained for both
the long and short versions of RBS clone 2, which indicates
that the additional insertion from the original EIPCR
10   reaction played no role in initiating translation.

It should be noted that DNA sequencing was used
to verify the recloned RBS sites in all the RBS clones
shown in Table 5.   All sequences were unchanged except for
clone 12, in which a thymidine nucleotide was deleted from
15   the sequence.   It is known from the literature that a
critical element within an RBS sequence is an area of three
or more bases in the mRNA that hybridize with the 3' end of
16S rRNA in bacteria (also known as the Shine and Dalgarno
sequence).   The underlined bases in Table 8 show potential
20   areas of mRNA complementarity to the 3' end of 16S rRNA.
Potential Shine and Dalgarno sequences were identified for
RBS clones 2 and 12, but not for clones 1 and 6.

It should be pointed out that the mutagenesis
process described here is very powerful due to its ability
25   to optimize sequences at both the RNA and the protein
level.   In the case of the RBS sequences, some clones were
isolated with improved expression levels, even though the
area of the RBS sequence contained a 5' unexpected
insertion of 17 nucleotides derived from an uncut primer
30   sequence.   In fact, these selected RBS sequences proved
themselves to function well, whether or not the accidental
insertion was present downstream.   In addition, some of
these sequences are working efficiently even though they do
not contain a potential Shine and Dalgarno type sequence.
35   Thus, this data illustrates the ability of the library

SUBSTITUTE SHEET (RULE 26)

-75-

approach to randomly incorporate sequences, both in
predicted locations and in unpredicted locations and to
assess the effectiveness of the mutations on protein
expression irrespective of the location of the mutations.

5

-76-

Table 8

DNA Sequences from Original Ribosome Binding Site Library

Clones

5   Sequence Names   DNA Sequence (5'-3' ton strand)*

    Omp A RBS   TAACGAGGCGCAAAAA

    (SEQ ID NO: 34)

    Library RBS

10   #1 long   AGACACGTACAAACCAATGAAAGAGACCTATTTACTC

    (SEQ ID NO: 35)

        #2 long   AAGCAAAGTCCCGCGAATGAAAGAGACCTATTT

    (SEQ ID NO: 36)

15

        #8 long   ACGTTTAAACAGACACATGAAAGAGACCTATTT

    (SEQ ID NO: 37)

        #12   GGAACTCAAAGGCACC

20 (SEQ ID NO: 38)

    _____

    *   DNA sequence is from the XBAl site (-16 position
                relative to the ompA signal peptide sequence) to
                the start of the Omp A signal sequence (-1
25   position).

                Bolded nucleotides indicate the DNA sequence
                corresponding to an insertional error due to
                blunt ended ligation of a EIPCR DNA primer that
30   failed to be cut by BSA1.

                Underlined nucleotides indicate areas of
                potential complementarity with the 3' end of *E.
                coli* 16S ribosomal RNA. (ie. potential Shine
35   Dalgarno sequences).

SUBSTITUTE SHEET (RULE 26)

-77-

Example 4
Creation of Novel Regulatory Sequences in Mammalian Cells
As an example of the application of this
5   invention to the development of novel regulatory sequences
in mammalian cells, this example provides a method for
producing signal peptide libraries in mammalian cells for
improved protein expression. Figure 7a illustrates the
vector contemplated for use in this invention, pGCEMK. The
10   vector uses the immunoglobulin Kappa Light Chain promoter
(from the kappa light chain gene of the murine monoclonal
antibody CEM 231) to direct transcription of a desired
gene. In this example it is driving the native CEM kappa
light chain variable region, but one skilled in the art
15   will be aware that any natural or recombinant variable
region could replace the native Variable region and the
same process of signal sequence mutagenesis could be
accomplished. Further, the entire light chain coding
sequences could be replaced by those from any eukaryotic
20   gene and a similar process carried out. In the present
example, the aim of the mutagenesis is to determine if an
alternate signal sequence can be obtained of equivalent or
improved efficacy, as measured by expression of mature
light chain.
25   Other components of the vector, pGCEMK, include
the major intron from the human kappa light chain gene,
containing its native enhancer sequence; the human kappa
constant region, containing its native polyadenylation
signal; a Xanthine-guanine phosphoribosyl transferase gene
30   (gptR) under the control of an SV 40 promoter to permit
selection of stable transformed cells; the bacterial colE1
origin of replication, to permit high copy number plasmid
replication in _E. coli_. The plasmid, pGCEMK, was
constructed from the plasmids pHF-1 and pMLCE-10, which
35   have been deposited with the American Type Culture

SUBSTITUTE SHEET (RULE 26)

-78-

Collection as Accession numbers 67,637 and 67,639,
respectively. These plasmids and the construction pGCEMK
are described in U.S. patent application Serial No.
07/727,719, filed July 2, 1991, which is incorporated
5   herein by reference in its entirety.

In this example the native signal sequence is
removed. In one vector, this region is replaced with the
first two codons of the native signal to allow efficient
initiation of translation to take place. These are fused
10 directly to the first codon of the Variable region. Random
60-65 mer oligonucleotides are inserted into the vector by
PCR. One skilled in the art has the option of conserving
the signal intron in the resultant construct or deleting
it, as long as the control containing the native signal has
15 a comparable format. Following incorporation of the random
nucleotides into the eukaryotic vector, the vector is
introduced into suitable eukaryotic cells that are capable
of expressing protein from the expression vector (for
example, SP 2/0 cells, American Type Culture Collection,
20 Rockville, MD). In this example, electroporation is used
to mediate transfection; however, those skilled in the art
will be readily able to select other equally appropriate
methods for introducing their selected expression vector
into a suitable cell type. Positive cells are preferably
25 selected by growing the cells on media containing
hypoxanthine, mycophenolic acid and xanthine (HMAX).

At this point, one skilled in the art could
enrich for cells actually expressing protein at a
detectable level, for instance, by Fluorescence Activated
30 Flow Cytometry with Cell Sorting. In this example, cells
expressing a detectable level of protein are identified by
ELISA. Positive colonies are further quantitated for
protein expression by a quantitative ELISA technique as
compared with cells expressing protein under the control of
35 the native regulatory sequence, and using the vector

SUBSTITUTE SHEET (RULE 26)

-79-

without random oligonucleotides added as a negative
control. Cells expressing equal or greater levels of
protein as compared with cells containing the native
regulatory sequence are subjected to DNA sequencing, and
5   identified as optimal signal sequences for the particular
protein expressed.

Example 5
10   Creation of Novel Protein Trafficking Signals in Mammalian
Cells
          A critical component of immunoglobulin (Ig)
protein trafficking is the binding of the chaperon protein
BiP to immunoglobulin heavy chain and the subsequent
15   displacement of BiP by light chain (Nature 306:387-389,
1983). Secretion of the intact immunoglobulin molecule is
not possible unless BiP is disassociated from the heavy
chain. The kinetics of these interactions impact directly
upon the level of expression of intact antibody. For
20   example, if binding of heavy chain monomers by BiP is
inefficient, non-bound heavy chain monomers can be
secreted. These monomers compete with intact molecules and
inhibit the secretion process. Alternatively, the
inefficient displacement of BiP, by light chain, from BiP-
25   associated heavy chain would slow the process of intact
immunoglobulin secretion. The BiP binding regions within
the immunoglobulin heavy chain have been narrowed to
specific regions on the molecule (in C_{HI} and V_H; Blood
79:2181-2195, 1992), encoded by specific regions of DNA in
30   the gene. The methods of this invention can be used to
identify optimized sequences that increase overall
expression of intact Ig without increasing the level of
light or heavy chain monomers (or homodimers).
          A mammalian expression vector is selected that
35   contains a heavy chain gene. Figure 8 illustrates one

SUBSTITUTE SHEET (RULE 26)

-80-

heavy chain chimeric vector, pNCEMG1, that is useful for
practicing the trafficking methods of this invention. This
vector uses the Immunoglobulin heavy chain promoter to
transcribe the desired gene inserted into a cloning site.

5   The vector includes an immunoglobulin heavy chain
polyadenylation signal and includes a neomycin resistance
gene. The neomycin resistance gene is under the control of
the SV40 promoter, thereby permitting the selection of
stable transformed cells in media containing the antibiotic

10 G418. The bacterial colE1 origin of replication permits
high copy number plasmid replication in *E. coli*.
Construction of this plasmid from plasmids pHG1z and
pMHCE30, which have been deposited in the American Type
Culture Collection (Accession Nos. 67,638 and 67,640,

15 respectively) as described in pending U.S. patent
application Serial No. 07/727,319. In this example, the
native CEM immunoglobulin heavy chain V-region sequence is
incorporated into the vector, but one skilled in the art
will be aware that any natural or recombinant variable

20 region could replace the native variable region and the
same process of BiP binding region optimization could be
carried out.

        The normal Bip binding region located within the
human gamma 1 constant region gene is operably linked to

25 the recombinant variable region in this vector. A second
vector is now constructed that is identical to the first
except that the BiP binding region is deleted from the C_{HI}
domain. Random 12-20 mer sequences are inserted into the
vector using the PCR methods described in Example 1.

30 Following incorporation of the random oligonucleotides into
the eukaryotic vector, the vector is introduced into
suitable eukaryotic cells, which produce a light chain
complementary to the heavy chain used (e.g. the cell lines
produced in Example 4). The vector can be introduced into

35 cells using methods well known in the art. Here,

SUBSTITUTE SHEET (RULE 26)

-81-

electroporation is used to mediate transfection. Positive cells are preferably selected by growing the cells in media containing G418.

5   At this point, one skilled in the art could enrich for cells actually expressing protein at a detectable level, for instance, by Fluorescence Activated Flow Cytometry with Cell Sorting. In this example, cells expressing a detectable level of protein are identified by ELISA. Positive cells are further quantitated for protein

10   expression by a quantitative ELISA technique as compared with cells expressing the protein under the control of the native regulatory sequence, and using the vector without random oligonucleotides added as a negative control. Cells expressing equal or greater levels of intact antibody as

15   compared with positive control cells and that do not express an increased level of free light or heavy chains are subjected to DNA sequencing to determine the DNA sequence of the trafficking signals most efficacious for expression of the immunoglobulin.

20   While particular embodiments of the invention have been described in detail, it will be apparent to those skilled in the art that these embodiments are exemplary rather than limiting, and the true scope of the invention is that defined in the following claims.

25

SUBSTITUTE SHEET (RULE 26)

-82-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5   (i)   APPLICANT: Antelman, Douglas E.
                                   Wilson, Barry S.

        (ii)   TITLE OF INVENTION: METHOD FOR CREATING
                  OPTIMIZED REGULATORY REGIONS AFFECTING
10   PROTEIN EXPRESSION AND PROTEIN TRAFFICKING

        (iii)   NUMBER OF SEQUENCES: 43

        (iv)   CORRESPONDENCE ADDRESS:
15   (A) ADDRESSEE: Eli Lilly and Company
                  (B) STREET: Lilly Corporate Center
                  (C) CITY: Indianapolis
                  (D) STATE: Indiana
                  (E) COUNTRY: United States of America
20   (F) ZIP: 46285

        (v)   COMPUTER READABLE FORM:
                  (A) MEDIUM TYPE: Floppy disk
                  (B) COMPUTER: Apple MacIntosh
25   (C) OPERATING SYSTEM: Apple System 7.0
                  (D) SOFTWARE: Microsoft Word

        (vi)   CURRENT APPLICATION DATA:
                  (A) APPLICATION NUMBER:
                  (B) FILING DATE:
30   (C) CLASSIFICATION:

        (viii)   ATTORNEY/AGENT INFORMATION:
                  (A) NAME: Gaylo, Paul J.
35   (B) REGISTRATION NUMBER: 36,808

SUBSTITUTE SHEET (RULE 26)

-83-

                    (C)   REFERENCE/DOCKET NUMBER: H-8589

          (ix)   TELECOMMUNICATION INFORMATION:
                    (A)   TELEPHONE: (317) 276-0756
   5   (B)   TELEFAX: (317) 276-3861

   (2)   INFORMATION FOR SEQ ID NO: 1:

          (i)   SEQUENCE CHARACTERISTICS:
   10   (A)   LENGTH: 26 base pairs
                    (B)   TYPE: nucleic acid
                    (C)   STRANDEDNESS: single
                    (D)   TOPOLOGY: linear

   15   (ii)   MOLECULE TYPE: cDNA

          (iii)   HYPOTHETICAL: NO

          (iv)   ANTI-SENSE: NO
   20

   25   (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 1:

   ATATATTCTA GATAACGAGG CGCAAA   26

   (2)   INFORMATION FOR SEQ ID NO: 2:
   30

          (i)   SEQUENCE CHARACTERISTICS:
                    (A)   LENGTH: 54 base pairs
                    (B)   TYPE: nucleic acid
                    (C)   STRANDEDNESS: single
   35   (D)   TOPOLOGY: linear

SUBSTITUTE SHEET (RULE 26)

-84-

(ii)   MOLECULE TYPE: cDNA

(iii)   HYPOTHETICAL: NO

5   (iv)   ANTI-SENSE: NO

10   (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ATAAATTGGT CTCACCAGAA CCACCACCAC CACCTGCAGA GACAGTGACC

AGAG   54

15   (2) INFORMATION FOR SEQ ID NO: 3:

(i)   SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 25 base pairs
                (B)   TYPE: nucleic acid
20   (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

(ii)   MOLECULE TYPE: cDNA

25   (iii)   HYPOTHETICAL: NO

(iv)   ANTI-SENSE: NO

30

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGTCTCGCGG TATCATTGCA GCACT   25

35

SUBSTITUTE SHEET (RULE 26)

-85-

(2)   INFORMATION FOR SEQ ID NO: 4:

        (i)   SEQUENCE CHARACTERISTICS:
                   (A)   LENGTH: 35 base pairs
   5   (B)   TYPE: nucleic acid
                   (C)   STRANDEDNESS: single
                   (D)   TOPOLOGY: linear

        (ii)   MOLECULE TYPE: cDNA
  10
        (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO

  15

        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 4:

  20   AATTGGTCTC GGAACCACGC TCACCGGCTC CAGAT   35

(2)   INFORMATION FOR SEQ ID NO: 5:

        (i)   SEQUENCE CHARACTERISTICS:
  25   (A)   LENGTH: 53 base pairs
                   (B)   TYPE: nucleic acid
                   (C)   STRANDEDNESS: single
                   (D)   TOPOLOGY: linear

  30   (ii)   MOLECULE TYPE: cDNA

        (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO
  35

SUBSTITUTE SHEET (RULE 26)

-86-

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 5:

5 ATAAATTGGT CTCATCTGGT GGTGGTGGTT CTCACCCAGA AACGCTGGTG

   AAA   53

   (2) INFORMATION FOR SEQ ID NO: 6:

10   (i)   SEQUENCE CHARACTERISTICS:
                     (A) LENGTH: 44 base pairs
                     (B) TYPE: nucleic acid
                     (C) STRANDEDNESS: single
                     (D) TOPOLOGY: linear

15
          (ii)   MOLECULE TYPE: cDNA

          (iii)   HYPOTHETICAL: NO

20   (iv)   ANTI-SENSE: NO

25
          (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 6:

   ATATATGGAT CCGGTACCCT ATCATTACCA ATGCTTAATC AGTG   44

30 (2) INFORMATION FOR SEQ ID NO: 7:

          (i)   SEQUENCE CHARACTERISTICS:
                     (A) LENGTH: 32 base pairs
                     (B) TYPE: nucleic acid
35   (C) STRANDEDNESS: single

SUBSTITUTE SHEET (RULE 26)

-87-

(D)   TOPOLOGY: linear

(ii)   MOLECULE TYPE: cDNA

5   (iii)   HYPOTHETICAL: NO

(iv)   ANTI-SENSE: NO

10

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATAATTAGGT CTCCCAGGCT GTGGTGACCC AG   32

15

(2)   INFORMATION FOR SEQ ID NO: 8:

(i)   SEQUENCE CHARACTERISTICS:
(A)   LENGTH: 32 base pairs
20   (B)   TYPE: nucleic acid
(C)   STRANDEDNESS: single
(D)   TOPOLOGY: linear

(ii)   MOLECULE TYPE: cDNA
25
(iii)   HYPOTHETICAL: NO

(iv)   ANTI-SENSE: NO

30

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 8:

35   ATAATTAGGT CTCACCTGTT TTTGCGCCTC GT   32

SUBSTITUTE SHEET (RULE 26)

-88-

(2)   INFORMATION FOR SEQ ID NO: 9:

     (i)   SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 31 base pairs
5
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: single
             (D)   TOPOLOGY: linear

     (ii)   MOLECULE TYPE: cDNA
10
     (iii)   HYPOTHETICAL: NO

     (iv)   ANTI-SENSE: NO

15

     (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 9:

20   ATAATTAGGT CTCAAGGCTG TGGTGACCCA G   31

(2)   INFORMATION FOR SEQ ID NO: 10:

     (i)   SEQUENCE CHARACTERISTICS:
25
             (A)   LENGTH: 100 base pairs
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: single
             (D)   TOPOLOGY: linear

30   (ii)   MOLECULE TYPE: cDNA

     (iii)   HYPOTHETICAL: NO

     (iv)   ANTI-SENSE: NO
35

-89-

    (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 10:

5   AATAATAGGT CTCTGCCTGN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    NNNNNNNNNN   60

    NNNNNNNNNN NNNNNNNNNN NNCATTTTTT GCGCCTCGTT   100

10 (2) INFORMATION FOR SEQ ID NO: 11:

        (i)   SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 21 amino acids
            (B)   TYPE: amino acid
15   (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

        (ii)   MOLECULE TYPE: peptide

20   (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO

        (v)   FRAGMENT TYPE: N-terminal
25

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
30
    Met Lys Lys Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe
    Ala
    1   5   10   15

SUBSTITUTE SHEET (RULE 26)

-90-

Thr Val Ala Gln Ala
                    20

(2)   INFORMATION FOR SEQ ID NO: 12:

    (i)   SEQUENCE CHARACTERISTICS:
               (A)   LENGTH: 22 amino acids
               (B)   TYPE: amino acid
               (C)   STRANDEDNESS: single
               (D)   TOPOLOGY: linear

    (ii)   MOLECULE TYPE: peptide

    (iii)   HYPOTHETICAL: NO

    (iv)   ANTI-SENSE: NO

    (v)   FRAGMENT TYPE: N-terminal

    (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Leu Tyr Pro Gly Glu Met Leu Cys Met Gln Asn Phe Ile Thr
Leu
1   5   10   15

Ile Ile Glu Ile Tyr Ile
                    20

(2)   INFORMATION FOR SEQ ID NO: 13:

    (i)   SEQUENCE CHARACTERISTICS:
               (A)   LENGTH: 22 amino acids

-91-

        (B)   TYPE: amino acid

        (C)   STRANDEDNESS: single

        (D)   TOPOLOGY: linear

5   (ii)   MOLECULE TYPE: peptide

     (iii)   HYPOTHETICAL: NO

      (iv)   ANTI-SENSE: NO

10

       (v)   FRAGMENT TYPE: N-terminal

15

      (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Gly Met Tyr Lys Leu Arg Leu Leu Ile Asp Cys Leu Met Leu Tyr

20 1   5   10   15

Val Cys Leu Thr Asp Thr

              20

25 (2)   INFORMATION FOR SEQ ID NO: 14:

      (i)   SEQUENCE CHARACTERISTICS:

        (A)   LENGTH: 21 amino acids

        (B)   TYPE: amino acid

30   (C)   STRANDEDNESS: single

        (D)   TOPOLOGY: linear

      (ii)   MOLECULE TYPE: peptide

35   (iii)   HYPOTHETICAL: NO

SUBSTITUTE SHEET (RULE 26)

-92-

(iv)   ANTI-SENSE: NO

(v)   FRAGMENT TYPE: N-terminal

(ix)   FEATURE:
          (A)   NAME/KEY: Stop Codon
          (B)   LOCATION: 8, 10

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Asn Ile Leu Leu Glu Asn Xaa Met Xaa Arg Thr His Asp Met
Leu
1   5   10   15

Ile Asp Gly Thr Ile
                20

(2) INFORMATION FOR SEQ ID NO: 15:

          (i)   SEQUENCE CHARACTERISTICS:
                    (A)   LENGTH: 21 amino acids
                    (B)   TYPE: amino acid
                    (C)   STRANDEDNESS: single
                    (D)   TOPOLOGY: linear

          (ii)   MOLECULE TYPE: peptide

          (iii)   HYPOTHETICAL: NO

          (iv)   ANTI-SENSE: NO

          (v)   FRAGMENT TYPE: N-terminal

SUBSTITUTE SHEET (RULE 26)

-93-

            (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 15:

  5 Met Gly Met Tyr Lys Leu Arg Leu Leu Asp Cys Leu Met Leu Tyr
     Val
     1   5   10   15

     Cys Ser Thr Asp Thr
 10   20

     (2) INFORMATION FOR SEQ ID NO: 16:

           (i)   SEQUENCE CHARACTERISTICS:
 15   (A)   LENGTH: 22 amino acids
                       (B)   TYPE: amino acid
                       (C)   STRANDEDNESS: single
                       (D)   TOPOLOGY: linear

 20   (ii)   MOLECULE TYPE: peptide

           (iii)   HYPOTHETICAL: NO

           (iv)   ANTI-SENSE: NO
 25
           (v)   FRAGMENT TYPE: N-terminal

 30
           (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 16:

     Met Thr Ser Val Cys Lys Ile Thr Val Leu Ile Leu Phe Xaa Thr
     Gly
 35 1   5   10   15

SUBSTITUTE SHEET (RULE 26)

-94-

Thr Tyr Asn Ile Leu Gly
                    20

(2)   INFORMATION FOR SEQ ID NO: 17:

       (i)   SEQUENCE CHARACTERISTICS:
                   (A)   LENGTH: 22 amino acids
                   (B)   TYPE: amino acid
                   (C)   STRANDEDNESS: single
                   (D)   TOPOLOGY: linear

       (ii)   MOLECULE TYPE: peptide

       (iii)   HYPOTHETICAL: NO

       (iv)   ANTI-SENSE: NO

       (v)   FRAGMENT TYPE: N-terminal

       (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Met Arg Asn Thr Tyr Val Tyr Xaa Lys Tyr Gly Arg Thr Glu Trp
Leu
1   5   10   15

Val Glu Gln Ile Asn Leu
                    20

(2)   INFORMATION FOR SEQ ID NO: 18:

       (i)   SEQUENCE CHARACTERISTICS:
                   (A)   LENGTH: 22 amino acids
                   (B)   TYPE: amino acid

-95-

         (C)   STRANDEDNESS: single
         (D)   TOPOLOGY: linear

    (ii)   MOLECULE TYPE: peptide

    (iii)   HYPOTHETICAL: NO

    (iv)   ANTI-SENSE: NO

    (v)   FRAGMENT TYPE: N-terminal

    (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Val Gln Cys Glu Arg Xaa Ile Tyr His Tyr Val Gln Arg Ile
Arg
1   5   10   15

Glu Tyr Phe Xaa Phe Glu
                20

(2)   INFORMATION FOR SEQ ID NO: 19:

    (i)   SEQUENCE CHARACTERISTICS:
         (A)   LENGTH: 22 amino acids
         (B)   TYPE: amino acid
         (C)   STRANDEDNESS: single
         (D)   TOPOLOGY: linear

    (ii)   MOLECULE TYPE: peptide

    (iii)   HYPOTHETICAL: NO

-96-

          (iv)   ANTI-SENSE: NO

          (v)   FRAGMENT TYPE: N-terminal

5   (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 19:

    Met Val Gly Ile Cys Thr Gly Ile Gly Val Gly Ile Leu Leu Leu
    Leu
    1   5   10   15
10
    Leu Val Val Gly Gln Val
                    20

    (2) INFORMATION FOR SEQ ID NO: 20:
15
          (i)   SEQUENCE CHARACTERISTICS:
                    (A)   LENGTH: 22 amino acids
                    (B)   TYPE: amino acid
                    (C)   STRANDEDNESS: single
20   (D)   TOPOLOGY: linear

          (ii)   MOLECULE TYPE: peptide

          (iii)   HYPOTHETICAL: NO
25
          (iv)   ANTI-SENSE: NO

          (v)   FRAGMENT TYPE: N-terminal

30

          (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 20:

SUBSTITUTE SHEET (RULE 26)

-97-

Met Leu Gly Ala Met Glu Phe Lys Lys Leu Ser Tyr Ala Ile Leu
Val
1   5   10   15

5   Leu Ser His Thr Pro Thr
                    20

(2) INFORMATION FOR SEQ ID NO: 21:

10   (i)   SEQUENCE CHARACTERISTICS:
                     (A)   LENGTH: 22 amino acids
                     (B)   TYPE: amino acid
                     (C)   STRANDEDNESS: single
                     (D)   TOPOLOGY: linear
15
          (ii) MOLECULE TYPE: peptide

          (iii)   HYPOTHETICAL: NO

20   (iv)   ANTI-SENSE: NO

          (v)   FRAGMENT TYPE: N-terminal

25

          (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 21:

     Met Asp Gln Gly Ile Lys Tyr Arg Leu Glu Lys Trp Arg Gly Arg
30   Ile
     1   5   10   15

     Ser Leu Ile Arg Val Phe
                    20
35

SUBSTITUTE SHEET (RULE 26)

-98-

(2)   INFORMATION FOR SEQ ID NO: 22:

        (i)   SEQUENCE CHARACTERISTICS:
                  (A)   LENGTH: 21 amino acids
                  (B)   TYPE: amino acid
                  (C)   STRANDEDNESS: single
                  (D)   TOPOLOGY: linear

        (ii)   MOLECULE TYPE: peptide

        (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO

        (v)   FRAGMENT TYPE: N-terminal

        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Met Gln Arg Arg Gly Ile Arg Trp Ala Tyr Val Ile Tyr Leu Ser
Met
1   5   10   15

Phe Phe Gly Ala Cys
                20

(2)   INFORMATION FOR SEQ ID NO: 23:

        (i)   SEQUENCE CHARACTERISTICS:
                  (A)   LENGTH: 22 amino acids
                  (B)   TYPE: amino acid
                  (C)   STRANDEDNESS: single
                  (D)   TOPOLOGY: linear

SUBSTITUTE SHEET (RULE 26)

-99-

(ii)   MOLECULE TYPE: peptide

(iii)   HYPOTHETICAL: NO

(iv)   ANTI-SENSE: NO

(v)   FRAGMENT TYPE: N-terminal

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 23:

Met Leu Gly Val Leu Thr Met Pro Cys Asp Thr Met Val Glu Asp
Leu
1   5   10   15

Ala Cys Asp Tyr Met Ser
                20

(2) INFORMATION FOR SEQ ID NO: 24:

(i)   SEQUENCE CHARACTERISTICS:
          (A)   LENGTH: 120 base pairs
          (B)   TYPE: nucleic acid
          (C)   STRANDEDNESS: single
          (D)   TOPOLOGY: linear

(ii)   MOLECULE TYPE: cDNA

(iii)   HYPOTHETICAL: NO

(iv)   ANTI-SENSE: NO

SUBSTITUTE SHEET (RULE 26)

-100-

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 24:

TAATTATTCT AGAATGGGGA TGTACAAGCT GAGACTACTG ATAGATTGCC

5   TAATGCTTTA   .60

TGTTTGTTCA ACCGATACCG GAATTCCGGG TCACCCAGAA ACGCTGGTGA

AAGTAAAAGA   120

(2)   INFORMATION FOR SEQ ID NO: 25:

10

        (i)   .   SEQUENCE CHARACTERISTICS:
                   (A)   LENGTH: 120 base pairs
                   (B)   TYPE: nucleic acid
                   (C)   STRANDEDNESS: single
15   (D)   TOPOLOGY: linear

        (ii)   MOLECULE TYPE: cDNA

        (iii)   HYPOTHETICAL: NO

20

        ·(iv)   ANTI-SENSE: NO

25

        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 25:

TAATTATTCT AGAATGACTT CTGGGAGCAA AATAACTGTT CTTATACTTC

TTTTGACTGG   60

30 TACATACAAT ATATTAGGGG GAATTCCGGG TCACCCAGAA ACGTCGGTGA

AAGTAAAAGA   120

(2)   INFORMATION FOR SEQ ID NO: 26:

35   (i)   SEQUENCE CHARACTERISTICS:

SUBSTITUTE SHEET (RULE 26)

-101-

                    (A)   LENGTH: 31 base pairs
                    (B)   TYPE: nucleic acid
                    (C)   STRANDEDNESS: single
                    (D)   TOPOLOGY: linear

        (ii)   MOLECULE TYPE: cDNA

        (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO

        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 26:

ATAATAGGTC TCCACCCAGA AACGCTGGTG A   31

    (2) INFORMATION FOR SEQ ID NO: 27:

        (i)   SEQUENCE CHARACTERISTICS:
                    (A)   LENGTH: 72 base pairs
                    (B)   TYPE: nucleic acid
                    (C)   STRANDEDNESS: single
                    (D)   TOPOLOGY: linear

        (ii)   MOLECULE TYPE: cDNA

        (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO

SUBSTITUTE SHEET (RULE 26)

-102-

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 27:

ATAATAGGTC TCTGGGTGNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

NCATTTTTTG   60

5   CGCCTCGTTA TC   72

(2)   INFORMATION FOR SEQ ID NO: 28:

(i)   SEQUENCE CHARACTERISTICS:

10   (A)   LENGTH: 84 base pairs

(B)   TYPE: nucleic acid

(C)   STRANDEDNESS: single

(D)   TOPOLOGY: linear

15   (ii)   MOLECULE TYPE: cDNA

(iii)   HYPOTHETICAL: NO

(iv)   ANTI-SENSE: NO

20

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 28:

25   ATAATAGGTC TCTGGGTGNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

NNNNNNNNNN   60

NNNCATTTTT TGCGCCTCGT TATC   84

(2)   INFORMATION FOR SEQ ID NO: 29:

30

(i)   SEQUENCE CHARACTERISTICS:

(A)   LENGTH: 96 base pairs

(B)   TYPE: nucleic acid

(C)   STRANDEDNESS: single

35   (D)   TOPOLOGY: linear

SUBSTITUTE SHEET (RULE 26)

-103-

(ii)   MOLECULE TYPE: cDNA

(iii)   HYPOTHETICAL: NO

5   (iv)   ANTI-SENSE: NO

10   (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 29:

ATAATAGGTC TCTGGGTGNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

NNNNNNNNNN   60

NNNNNNNNNN NNNNNCATTT TTTGCGCCTC GTTATC   96

15

(2) INFORMATION FOR SEQ ID NO: 30:

        (i)   SEQUENCE CHARACTERISTICS:
                   (A) LENGTH: 31 base pairs
20   (B) TYPE: nucleic acid
                   (C) STRANDEDNESS: single
                   (D) TOPOLOGY: linear

        (ii)   MOLECULE TYPE: cDNA
25

        (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO

30

        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 30:

35   ATAAATAGGT CTCATGAAAA AGACAGCTAT C   31

SUBSTITUTE SHEET (RULE 26)

-104-

(2)   INFORMATION FOR SEQ ID NO: 31:

    (i)   SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 31 base pairs
5   (B)   TYPE: nucleic acid
                (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

    (ii)   MOLECULE TYPE: cDNA
10

    (iii)   HYPOTHETICAL: NO

    (iv)   ANTI-SENSE: NO

15

    (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 31:

20   ATAAATAGGT CTCCTCTAGA AATTGTGAAA T   31

(2)   INFORMATION FOR SEQ ID NO: 32:

    (i)   SEQUENCE CHARACTERISTICS:
25   (A)   LENGTH: 32 base pairs
                (B)   TYPE: nucleic acid
                (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

30   (ii)   MOLECULE TYPE: cDNA

    (iii)   HYPOTHETICAL: NO

    (iv)   ANTI-SENSE: NO
35

SUBSTITUTE SHEET (RULE 26)

-105-

        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 32:

5   ATAAATAGGT CTCATGAAAA AGACAGCTAT CG   32

    (2)   INFORMATION FOR SEQ ID NO: 33:

        (i)   SEQUENCE CHARACTERISTICS:
10   (A) LENGTH: 53 base pairs
                   (B) TYPE: nucleic acid
                   (C) STRANDEDNESS: single
                   (D) TOPOLOGY: linear

15   (ii)   MOLECULE TYPE: cDNA

        (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO
20

        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 33:
25
    ATAAATAGGT CTCTTTCATN NNNNNNNNNN NNNNNTCTAG AAATGTGTGA
    AAT   53

    (2)   INFORMATION FOR SEQ ID NO: 34:
30
        (i)   SEQUENCE CHARACTERISTICS:
                   (A) LENGTH: 16 base pairs
                   (B) TYPE: nucleic acid
                   (C) STRANDEDNESS: single
35   (D) TOPOLOGY: linear

SUBSTITUTE SHEET (RULE 26)

-106-

(ii)   MOLECULE TYPE: cDNA

(iii)   HYPOTHETICAL: NO

5   (iv)   ANTI-SENSE: NO

10   (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 34:

TAACGAGGCG CAAAAA   16

(2) INFORMATION FOR SEQ ID NO: 35:
15
        (i)   SEQUENCE CHARACTERISTICS:
                   (A) LENGTH: 37 base pairs
                   (B) TYPE: nucleic acid
                   (C) STRANDEDNESS: single
20   (D) TOPOLOGY: linear

        (ii)   MOLECULE TYPE: cDNA

        (iii)   HYPOTHETICAL: NO
25
        (iv)   ANTI-SENSE: NO

30
        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 35:

AGACACGTAC AAACCAATGA AAGAGACCTA TTTACTC   37

SUBSTITUTE SHEET (RULE 26)

-107-

(2)   INFORMATION FOR SEQ ID NO: 36:

        (i)   SEQUENCE CHARACTERISTICS:
                 (A)   LENGTH: 33 base pairs
5   (B)   TYPE: nucleic acid
                 (C)   STRANDEDNESS: single
                 (D)   TOPOLOGY: linear

        (ii)   MOLECULE TYPE: cDNA
10
        (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO

15

        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 36:

20   AAGCAAAGTC CCGCGAATGA AAGAGACCTA TTT   33

     (2)   INFORMATION FOR SEQ ID NO: 37:

        (i)   SEQUENCE CHARACTERISTICS:
25   (A)   LENGTH: 33 base pairs
                 (B)   TYPE: nucleic acid
                 (C)   STRANDEDNESS: single
                 (D)   TOPOLOGY: linear

30   (ii)   MOLECULE TYPE: cDNA

        (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO
35

SUBSTITUTE SHEET (RULE 26)

-108-

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 37:

ACGTTTAAAC AGACACATGA AAGAGACCTA TTT   33

(2) INFORMATION FOR SEQ ID NO: 38:

       (i)   SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 16 base pairs
                (B) TYPE: nucleic acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

       (ii)   MOLECULE TYPE: cDNA

       (iii)   HYPOTHETICAL: NO

       (iv)   ANTI-SENSE: NO

       (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GGAACTCAAA GGCACC   16

(2) INFORMATION FOR SEQ ID NO: 39:

       (i)   SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 61 base pairs
                (B) TYPE: nucleic acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

SUBSTITUTE SHEET (RULE 26)

-109-

    (ii)   MOLECULE TYPE: cDNA

    (iii)   HYPOTHETICAL: NO

5   (iv)   ANTI-SENSE: NO

10   (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CACACATTTC TAGAAGACAC GTACAAACCA ATGAAAAAAG ACAGCTATCG
CGATTGCAGT   60
G   61

15

(2)   INFORMATION FOR SEQ ID NO:40:

    (i)   SEQUENCE CHARACTERISTICS:
              (A)   LENGTH: 87 base pairs
20   (B)   TYPE: nucleic acid
              (C)   STRANDEDNESS: single
              (D)   TOPOLOGY: linear

    (ii)   MOLECULE TYPE: cDNA
25
    (iii)   HYPOTHETICAL: NO

    (iv)   ANTI-SENSE: NO

30

    (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 40:

SUBSTITUTE SHEET (RULE 26)

-110-

CACATTTCTA GAAAGCAAAG TCCCGCGAAT GAAAGAGACC TATTTATGAA

AAAGACAGCT   60

ATCGCGATTG CAGTGGCACT GGCTGGT   87

5   (2)   INFORMATION FOR SEQ ID NO: 41:

        (i)   SEQUENCE CHARACTERISTICS:
                   (A)   LENGTH: 70 base pairs
                   (B)   TYPE: nucleic acid
10   (C)   STRANDEDNESS: single
                   (D)   TOPOLOGY: linear

        (ii)   MOLECULE TYPE: cDNA

15   (iii)   HYPOTHETICAL: NO

        (iv)   ANTI-SENSE: NO

20

        (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 41:

CACATTTCTA GAAAGCAAAG TCCCGCGAAT GAAAAAGACA GCTATCGCGA

25 TTGCAGTGGC   60

ACTGGCTGGT   70

        (2)   INFORMATION FOR SEQ ID NO: 42:

30
        (i)   SEQUENCE CHARACTERISTICS:
                   (A)   LENGTH: 70 base pairs
                   (B)   TYPE: nucleic acid
                   (C)   STRANDEDNESS: single
35   (D)   TOPOLOGY: linear

SUBSTITUTE SHEET (RULE 26)

-111-

(ii)   MOLECULE TYPE: cDNA

(iii)   HYPOTHETICAL: NO

5   (iv)   ANTI-SENSE: NO

10   (xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 42:

CACATTTCTA GAACGTTTAA ACAGACACAT GAAAAAGACA GCTATCGCGA

TTGCAGTGGC   60

15 ACTGGCTGGT   70

(2) INFORMATION FOR SEQ ID NO: 43:

(i)   SEQUENCE CHARACTERISTICS:
20   (A)   LENGTH: 70 base pairs
                (B)   TYPE: nucleic acid
                (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

25   (ii)   MOLECULE TYPE: cDNA

(iii)   HYPOTHETICAL: NO

(iv)   ANTI-SENSE: NO
30

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO: 43:
35

SUBSTITUTE SHEET (RULE 26)

-112-

CACATTTCTA GAGGAACTCA AAGGCACCAT GAAAAAGACA GCTATCGCGA

TTGCAGTGGC   60

ACTGGCTGGT   70

-113-

Claims:

        1.   A method for optimizing the production of
polypeptide in a cell, comprising the steps of:
5   (a)   identifying at least one regulatory region
within a nucleic acid sequence to be mutagenized;
             (b)   preparing a nucleic acid vector comprising
said regulatory region and a nucleic acid sequence encoding
at least one polypeptide regulated by said regulatory
10   region;
             (c)   deleting said regulatory region from said
vector;
             (d)   producing a pool of random oligonucleotides;
             (e)   using a polymerase chain reaction to
15   introduce at least one random oligonucleotide into the
position previously occupied by said regulatory region in a
plurality of vectors to generate a pool of mutagenized
vectors;
             (f)   introducing said mutagenized vectors into a
20   cell sample;
             (g)   assaying for the expression of said
polypeptide in said cell sample;
             (h)   selecting cells exhibiting optimized
polypeptide expression; and
25   (i)   isolating optimized polypeptide from the
cells of step (h).
        2.   The method of Claim 1, wherein said
regulatory region is located within said nucleic acid
sequence encoding polypeptide.
30   3.   The method of Claim 1, wherein said
regulatory region is in a translated portion of said
nucleic acid sequence encoding polypeptide.
        4.   The method of Claim 1 wherein said
regulatory region is selected from the group consisting of
35   a signal sequence, a ribosome binding site, a promoter

SUBSTITUTE SHEET (RULE 26)

-114-

sequence, a translational regulatory sequence, a
transcription regulatory sequence, and a protein
trafficking sequence.

5.   The method of Claim 1, wherein said nucleic

5   acid sequence encodes an antibiotic resistant gene and the
selection step additionally consists essentially of growing
said cell sample in the presence of an antibiotic.

6.   The method of Claim 1, wherein said nucleic
acid sequence encoding polypeptide encodes a selectable

10   marker.

7.   The method of Claim 1, wherein said nucleic
acid sequence encoding polypeptide encodes a fusion
protein.

8.   The method of Claim 1, wherein said nucleic

15   acid sequence encoding polypeptide is derived from a
eukaryotic cell and said cell samples is prokaryotic.

9.   The method of Claim 1, wherein said random
oligonucleotides are biased.

10.   The method of Claim 1, additionally

20   comprising the steps of introducing said vector comprising
said regulatory region and a nucleic acid sequence encoding
at least one polypeptide regulated by said regulatory
region into a second cell sample, expressing said
polypeptide encoded by said vector in said second cell

25   sample, and measuring the level of polypeptide expression
in said second cell sample.

11.   The method of Claim 10, additionally
comprising the step of selecting cells from said first cell
sample that exhibit optimized polypeptide expression

30   relative to the measuring step of Claim 10.

12.   The method of Claim 11, additionally
comprising the step of determining the nucleic acid
sequence of the mutagenized regulatory region producing
optimized polypeptide expression.

SUBSTITUTE SHEET (RULE 26)

-115-

13.   The method of Claim 11, wherein said optimized polypeptide expression is a level of polypeptide expression greater than or equal to the level of polypeptide expression obtained from the measuring step of
5   Claim 10.

14.   A method for creating and isolating novel signal sequences, comprising the steps of:

(a)   identifying a signal sequence region within a nucleic acid sequence encoding a polypeptide;

10   (b)   preparing a nucleic acid vector comprising said nucleic acid sequence encoding the polypeptide;

(c)   introducing said vector into a first cell sample and expressing said polypeptide in said first cell sample;

15   (d)   measuring the level of polypeptide expression in said first cell sample;

(e)   deleting said signal sequence region from said vector;

(f)   producing a pool of random oligonucleotides;

20   (g)   using a polymerase chain reaction to introduce at least one of the random oligonucleotides into the position previously occupied by said signal sequence in a plurality of vectors to generate a pool of mutagenized vectors;

25   (h)   introducing said mutagenized vectors into a second cell sample;

(i)   assaying for the expression of said polypeptide in said second cell sample;

(j)   selecting cells exhibiting optimized
30   polypeptide expression relative to step (d); and

(k)   determining the nucleic acid sequence of the regulatory region located in the mutagenized vector contained in the cells of Step (j).

15.   The method of Claim 14, wherein said signal
35   sequence region comprises an entire signal sequence.

SUBSTITUTE SHEET (RULE 26)

-116-

16. The method of Claim 14, wherein said nucleic acid vector encodes an antibiotic resistant gene and the selection step additionally consists essentially of growing said cell sample in the presence of an antibiotic.

5   17. The method of Claim 14, wherein said nucleic acid sequence encoding polypeptide encodes a selectable marker.

18. The method of Claim 14, wherein said polymerase chain reaction is the enzymatic inverse

10   polymerase chain reaction.

19. The method of Claim 14, wherein said nucleic acid sequence encoding polypeptide encodes a fusion protein.

20. The method of Claim 16, wherein a portion of

15   said fusion protein is derived from β-lactamase.

21. The method of Claim 20, wherein a portion of said fusion protein is derived from an antibody.

22. The method of Claim 15, wherein said nucleic acid sequence encoding polypeptide is derived from a

20   eukaryotic cell and said first and second cell samples are prokaryotic.

23. The method of Claim 22, wherein said cells are *Escherichia coli*.

24. The method of Claim 15, wherein said random

25   oligonucleotides are biased.

25. The method of Claim 24, wherein said random oligonucleotides contain at least one positively charged amino acid at the N-terminus, a stretch of at least 8 hydrophobic amino acids and a small amino acid such as

30   alanine, glycine, or valine positioned at the C-terminus.

26. The method of Claim 15, wherein said optimized level of polypeptide expression is a level greater than or equal to the level of expression of said polypeptide in step (d).

SUBSTITUTE SHEET (RULE 26)

-117-

27.   A method for optimizing polypeptide
expression in a cell by performing random mutagenesis on a
regulatory region regulating polypeptide expression from a
nucleic acid vector wherein the improvement comprises
5   deleting a region of nucleic acid to be mutagenized,
isolating said vector containing said deletion and
replacing said region with random nucleic acid sequences.

28.   A method for creating and isolating novel
ribosome binding sites, comprising the steps of:
10   (a)   identifying a ribosome binding site within a
nucleic acid sequence;
        (b)   preparing a nucleic acid vector comprising
said nucleic acid sequence containing a ribosome binding
site operably linked to a nucleic acid sequence encoding a
15   polypeptide;
        (c)   introducing said vector into a first cell
sample and expressing said polypeptide in said first cell
sample;
        (d)   measuring the level of polypeptide
20   expression in said first cell sample;
        (e)   deleting said nucleic acid sequence
containing said ribosome binding site from said vector;
        (f)   producing a pool of random oligonucleotides;
        (g)   using a polymerase chain reaction to
25   introduce at least one of the random oligonucleotides into
the position previously occupied by said nucleic acid
sequence containing said ribosome binding site in a
plurality of vectors to generate a pool of mutagenized
vectors;
30   (h)   introducing said mutagenized vectors into a
second cell sample;
        (i)   assaying for the expression of said
polypeptide in said second cell sample;
        (j)   selecting cells exhibiting optimized
35   polypeptide expression relative to step (d); and

SUBSTITUTE SHEET (RULE 26)

-118-

(k)   determining the nucleic acid sequence containing said ribosome binding site located in the mutagenized vector introduced into the cells of step (j).

29.   The method of Claim 28, wherein said nucleic
5   sequence encodes an antibiotic resistance gene and said selection step consists essentially of growing said cell samples in the presence of an antibiotic.

30.   The method of Claim 28, wherein said nucleic acid sequence encoding polypeptide encodes a selectable
10   marker.

31.   The method of Claim 30, wherein said nucleic acid sequence encoding polypeptide is a fusion protein.

32.   The method of Claim 30, wherein a portion of said fusion protein is derived from an antibody.

15   33.   The method of Claim 31, wherein a portion of said fusion protein is derived from β-lactamase.

34.   The method of Claim 28, wherein said nucleic acid sequence encoding polypeptide is derived from a eukaryotic cell and said first and second cell samples are
20   prokaryotic.

35.   The method of Claim 31, wherein said cells are *Escherichia coli*.

36.   The method of Claim 28, wherein said random oligonucleotides are biased.

25   37.   The method of Claim 28, wherein said polymerase chain reaction is the enzymatic inverse polymerase chain reaction.

38.   The method of Claim 28, wherein said optimized level of polypeptide expression is a level
30   greater than or equal to the level of expression of said polypeptide in step (d).

39.   A protein signal sequence corresponding to SEQ ID NO: 13 and to protein signal sequences containing at least contiguous amino acid 10-mers thereof.

SUBSTITUTE SHEET (RULE 26)

-119-

40.   A protein signal sequence corresponding to
SEQ ID NO: 16 and to protein signal sequences containing at
least contiguous amino acid 10-mers thereof.

41.   A protein signal sequence corresponding to
5   SEQ ID NO: 14 and to protein signal sequences containing at
least contiguous amino acid 10-mers thereof.

42.   Isolated nucleic acid containing a ribosome
binding site sequence identified as SEQ ID NO: 35 and to
ribosome binding sites containing at least contiguous 5-
10   mers thereof.

43.   Isolated nucleic acid containing a ribosome
binding site sequence identified as SEQ ID NO: 37 and to
ribosome binding sites containing at least contiguous 5-
mers thereof.

15   44.   Isolated nucleic acid containing a ribosome
binding site sequence identified as SEQ ID NO: 38 and to
ribosome binding sites containing at least contiguous 5-
mers thereof.

45.   A method for creating and isolating novel
20   signal sequences, comprising the steps of:

(a)   identifying a signal sequence region within
a nucleic acid sequence encoding a polypeptide;

(b)   preparing a nucleic acid vector comprising
said nucleic acid sequence encoding a polypeptide, wherein
25   said polypeptide is a fusion protein having a C-terminus
corresponding to β-lactamase;

(c)   deleting said signal sequence region from
said vector;

(d)   producing a pool of random oligonucleotides
30   suitable for an enzymatic inverse polymerase chain
reaction;

(e)   using an enzymatic inverse polymerase chain
reaction to introduce said random oligonucleotides into the
position in said vector previously occupied by said signal
35   sequence to generate a pool of mutagenized vectors;

SUBSTITUTE SHEET (RULE 26)

-120-

        (f)   introducing said mutagenized vectors into a
first cell sample of *E. coli*;
        (g)   assaying for the expression of said
polypeptide in said first cell sample in the presence of
5   ampicillin; and
        (h)   selecting cells exhibiting optimized
polypeptide expression.
        46. The method of Claim 45, wherein said signal
sequence region comprises an entire signal sequence.
10   47. The method of Claim 45, additionally
comprising the steps of introducing said vector comprising
said nucleic acid sequence encoding a polypeptide, wherein
said polypeptide is a fusion protein having a C-terminus
corresponding to β-lactamase into a second cell sample of,
15   *E. coli* expressing said polypeptide in said second cell
sample and measuring the level of polypeptide expression in
said second cell sample.
        48. The method of Claim 47, additionally
comprising the step of selecting cells from said first cell
20   sample that exhibit optimized polypeptide expression
relative to the level of polypeptide expression observed in
the measuring step of Claim 49.
        49. The method of Claim 48, wherein said
optimized polypeptide expression is a level of polypeptide
25   expression greater than or equal to the level of
polypeptide expression obtained from the measuring step of
Claim 47.
        50. The method of Claim 46 or 48 additionally
comprising the step of determining the nucleic acid
30   sequence of said regulatory region contained in the
mutagenized vector isolated from cells identified by said
selecting step.
        51. The method of Claim 46, wherein the
concentration of ampicillin in step (i) is at least 30
35   μg/ml.

SUBSTITUTE SHEET (RULE 26)

-121-

52. The method of Claim 46, wherein said nucleic acid sequence encoding a polypeptide encodes a single-chain antibody.

53. The method of Claim 46, wherein said random
5 oligonucleotides are biased.

54. The method of Claim 53, wherein said random oligonucleotides contain at least one positively charged amino acid at the N-terminus, a stretch of at least 8 hydrophobic amino acids and a small amino acid such as
10 alanine, glycine or valine positioned at the C-terminus.

55. A method for identifying a novel nucleic acid sequence encoding a protein trafficking signal that directs a polypeptide to a desired location in a cell, comprising the steps of:
15   (a)   identifying a region of nucleic acid containing at least one protein trafficking signal to be mutagenized;

        (b)   preparing a nucleic acid vector comprising said protein trafficking signal sequence to be mutagenized
20 and a nucleic acid sequence encoding at least one polypeptide;

        (c)   deleting said protein trafficking signal sequence from said vector;

        (d)   producing a pool of random oligonucleotides;
25   (e)   using a polymerase chain reaction to introduce at least one of the random oligonucleotides into the position previously occupied by said protein trafficking signal in a plurality of vectors to generate a pool of mutagenized vectors;
30   (f)   introducing said mutagenized vectors into a cell sample;

        (g)   assaying for the location of said polypeptide in said cell sample;

        (h)   selecting cells expressing said polypeptide
35 in the desired cell location; and

SUBSTITUTE SHEET (RULE 26)

-122-

(i) determining the nucleic acid sequence of the protein trafficking signal contained in the mutagenized vector located in the cells of Step (h).

56. The method of Claim 55, wherein said protein trafficking sequence to be mutagenized is located in said nucleic acid sequence encoding said polypeptide.

57. The method of Claim 55, wherein said polymerase chain reaction is the enzymatic inverse polymerase chain reaction.

58. The method of Claim 55, wherein said desired cell location is extracellular.

59. The method of Claim 55, wherein said nucleic sequence encodes an antibiotic resistant gene and said selection step consists essentially of growing said cell sample in the presence of an antibiotic.

60. The method of Claim 55, wherein said polypeptide is a selectable marker.

61. The method of Claim 55, wherein said polypeptide is a fusion protein.

62. The method of Claim 55, wherein said polypeptide is derived from a eukaryotic cell and said cell sample is prokaryotic.

63. The method of Claim 55, wherein said random oligonucleotides are biased.

64. A method for creating and isolating novel regulatory sequences for optimizing the expression of a recombinant polypeptide in prokaryotic cells comprising the steps of:

(a) identifying at least one regulatory region within a nucleic acid sequence to be mutagenized;

(b) preparing a nucleic acid vector suitable for expressing polypeptide in a prokaryotic cell, said nucleic acid vector comprising said regulatory region and a nucleic acid sequence encoding at least one polypeptide operably linked to said regulatory region;

SUBSTITUTE SHEET (RULE 26)

-123-

(c)   introducing said vector into a first prokaryotic cell sample and expressing said polypeptide encoded by said vector in said first cell sample;

(d)   measuring the level of polypeptide

5   expression in said first prokaryotic cell sample;

(e)   deleting said regulatory region from said vector;

(f)   producing a pool of random oligonucleotides;

(g)   using a polymerase chain reaction to

10   introduce at least one random oligonucleotide into said position previously occupied by said regulatory region in a plurality of vectors to generate a pool of mutagenized vectors;

(h)   introducing said mutagenized vectors into a

15   first prokaryotic cell sample;

(i)   assaying for the expression of the polypeptide in said second prokaryotic cell sample;

(j)   selecting cells exhibiting optimized polypeptide expression relative to step (d); and

20   (k)   determining the nucleic acid sequence of said mutagenized regulatory region located within the vector introduced into the cells of step (j).

65.   The method of Claim 64, wherein said regulatory region comprises an entire regulatory sequence.

25   66.   The method of Claim 65, wherein said regulatory region is selected from the group consisting of a signal sequence, a ribosome binding site, a promoter sequence, a translational regulatory sequence, a transcription regulatory sequence and a protein trafficking

30   sequence.

67.   The method of Claim 65, wherein said nucleic acid sequence encoding polypeptide encodes an antibiotic resistant gene and the selection step additionally consists of growing said second prokaryotic cell sample in the

35   presence of an antibiotic.

SUBSTITUTE SHEET (RULE 26)

-124-

68. The method of Claim 65, wherein said nucleic acid sequence encoding polypeptide encodes a selectable marker.

69. The method of Claim 65, wherein said nucleic acid sequence encoding polypeptide encodes a fusion protein.

70. The method of Claim 65, wherein said random oligonucleotides are biased.

71. The method of Claim 65, wherein said nucleic acid sequence encoding polypeptide encodes a polypeptide derived from a eukaryotic cell.

72. The method of Claim 65, wherein said first and second cell samples are derived from the same cell type.

B456

5' ATATATTCTAGATAACGAGGCGCAAA
Xbal
B444

B456
3' GAGACCAGTGACAGAGAGGTCCACCACCACCAAGACCA CTCTGGTTAAATA 5'
gly gly gly gly gly ser   Bsa I

V_H

V_L

pUCHASC2

TEM β-lactamase

B457
ATAAATTGGTCTCATCTGGTGGTGGTTCTCACCCAGAAACGCTGGTGAAA
Bsa I   ser gly gly gly ser his pro glu thr leu val lys

B447
GTGACTAATTCGTAACCATTACTATCCCATGGCCTAGGTATATA
stp stp stp   Kpnl   BamHI

FIG. I A

FIG. I B

FIG.IC

4 / 22

FIG.2 A

FIG.2 B

FIG. 3

Wildtype OMPA

M K K T A I A I A V A L A G F A T V A Q A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Protein Sequence

- Alpha, Regions - Garnier-Robson
- Alpha, Regions - Chou-Fasman
- Beta, Regions - Garnier-Robson
- Beta, Regions - Chou-Fasman

- Hydrophilicity Plot - Kyte-Doolittle

FIG. 4 A

E-10

Protein Sequence

■ Alpha, Regions - Garnier-Robson
■ Alpha, Regions - Chou-Fasman
■ Beta, Regions - Garnier-Robson
■ Beta, Regions - Chou-Fasman

■ Hydrophilicity Plot - Kyte-Doolittle

M T S G S K I T V L I L L L T G T Y N I L G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

FIG. 4B

E-5

FIG. 4C

FIG. 4D

FIG. 4E

LAMB

M M I T L R K L P L A V A V A A G V M S A Q A M A

Protein Sequence

- ■ Alpha, Regions - Garnier-Robson
- ■ Alpha, Regions - Chou-Fasman
- ■ Beta, Regions - Garnier-Robson
- ■ Beta, Regions - Chou-Fasman

- ■ Hydrophilicity Plot - Kyte-Doolittle

FIG. 4F

Bsa I

5' ATAATAGGTCTC CACCCAGAAACGCTGGTGA

B755

Bsa I

3' CTATTGCTCCGCGTTTTTTAC (N)₃₃ GTGGGTCTCTGG ATAATA 5'
3' CTATTGCTCCGCGTTTTTTAC (N)₄₅ GTGGGTCTCTGG ATAATA 5'
3' CTATTGCTCCGCGTTTTTTAC (N)₅₇ GTGGGTCTCTGG ATAATA 5'

B756
B757
B758

lac

V_L   V_H   amp R

cam R

pCCHAsc1-ampRv2ΔSP
3.155 kB

FIG. 5A

FIG. 5B

pC AmpR SP-12/16/20
2.441/ 2.453/ 2.465 kB

Plasmid Libraries containing random leader peptide of 12, 16 or 20 amino acids. The random signal peptides are used to direct ß-lactamase to the periplasm of E. coli
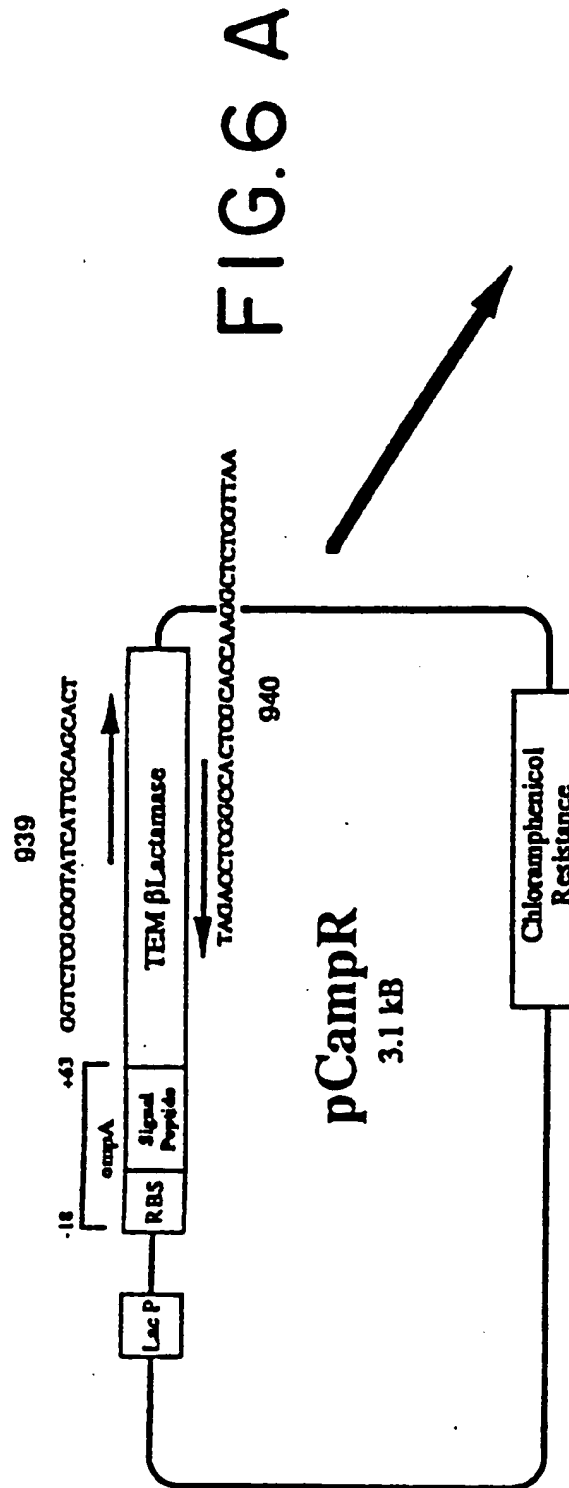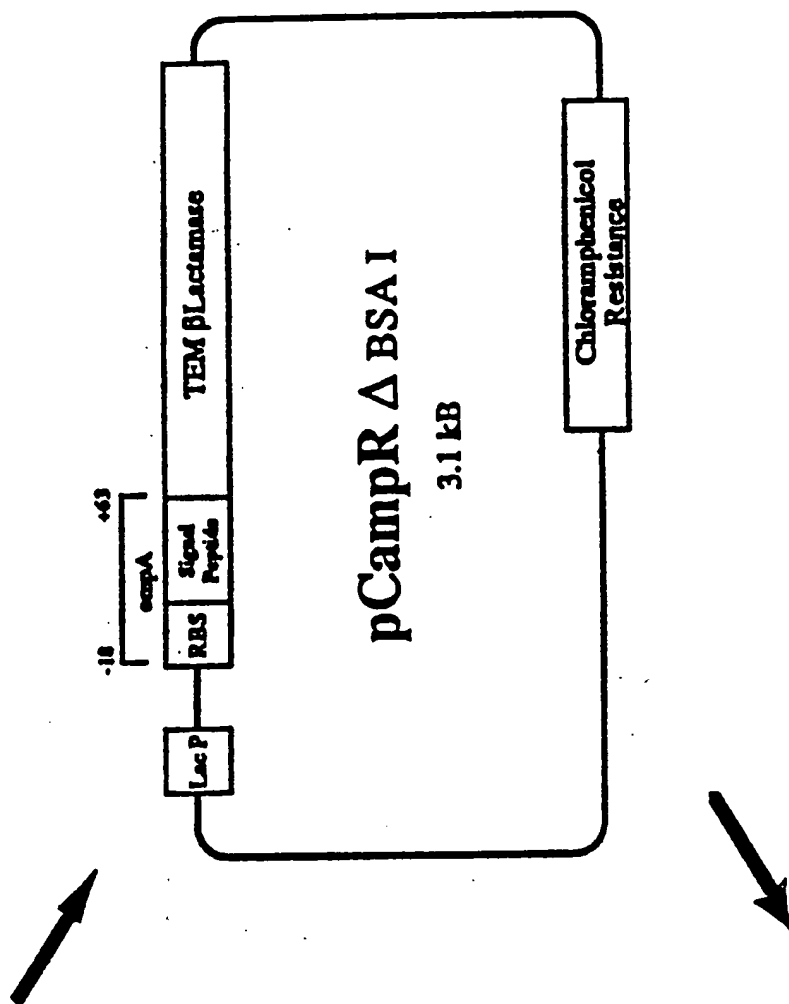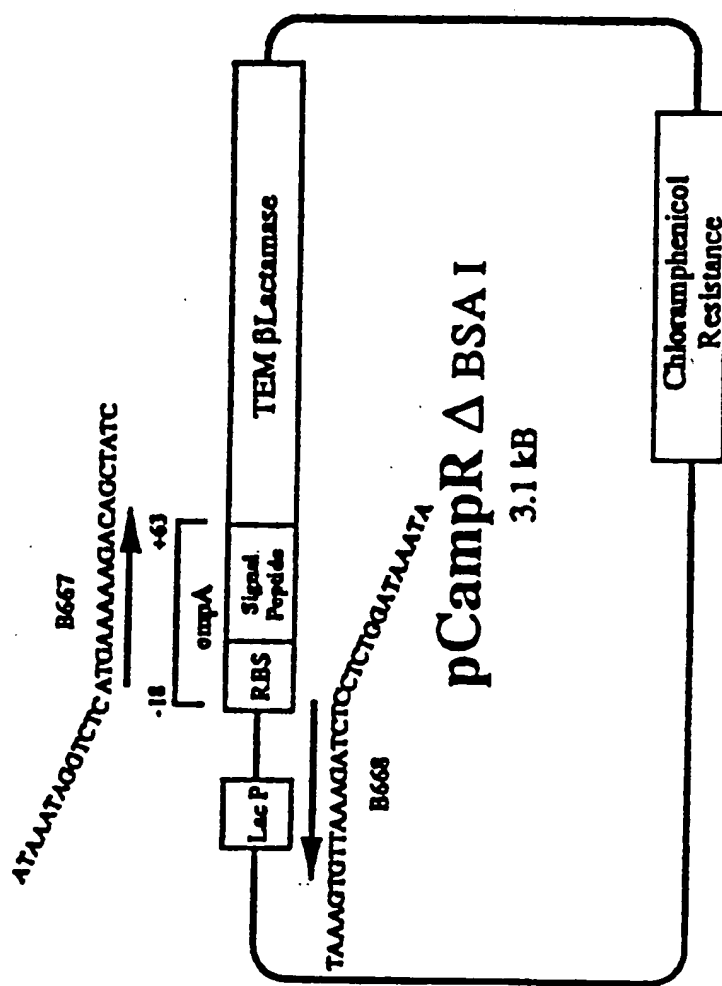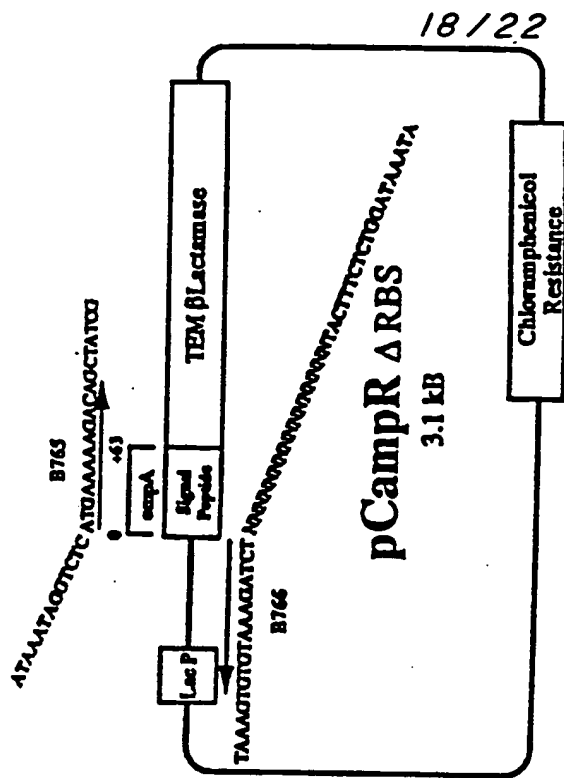
FIG.6 A

FIG.6B

17/22

FIG.6C

18/22

FIG.6D
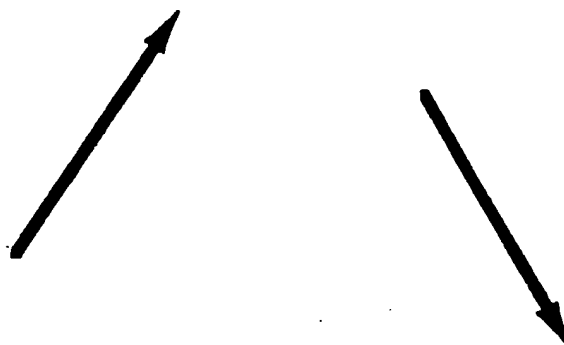
19 / 22
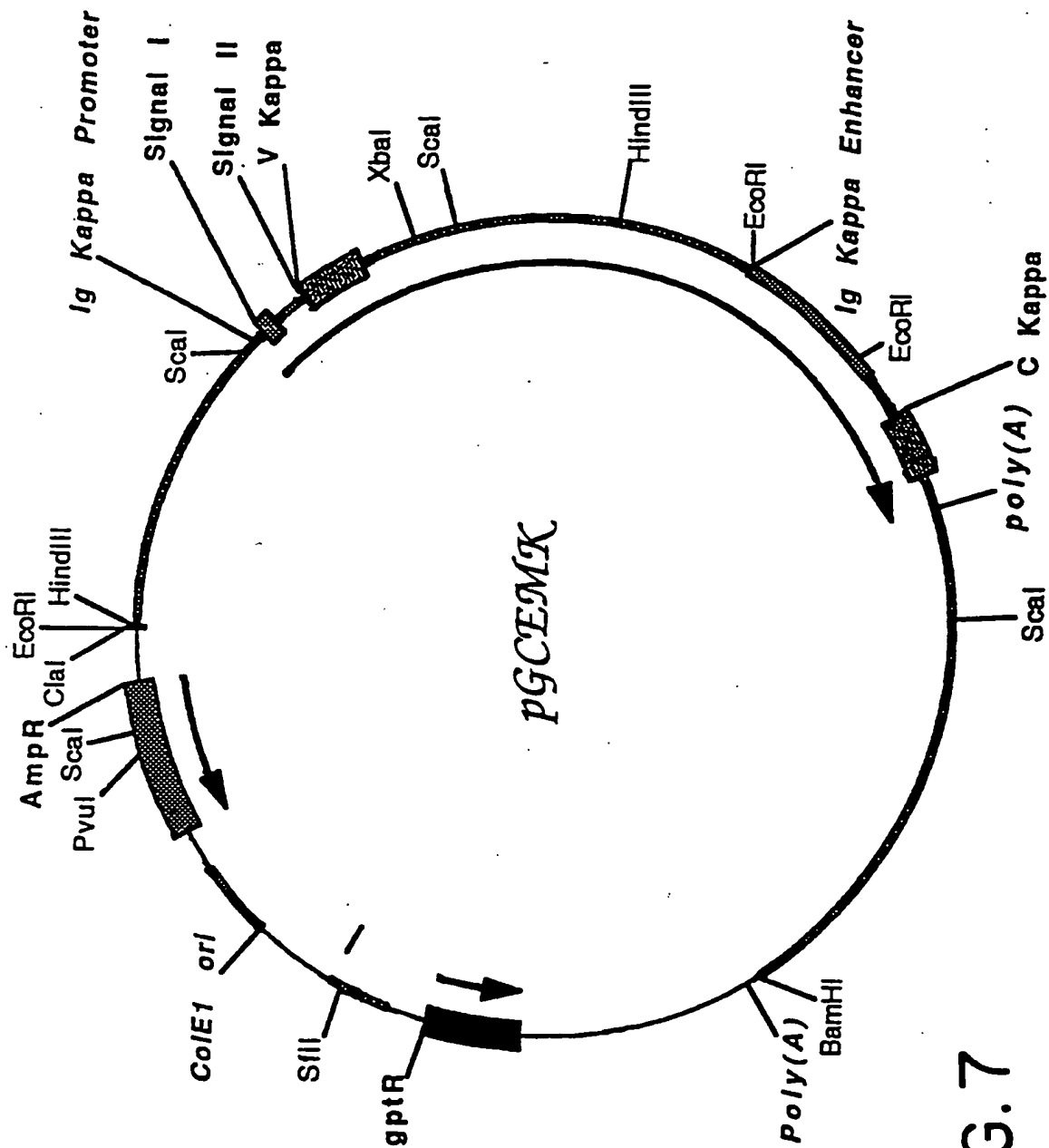
FIG.7
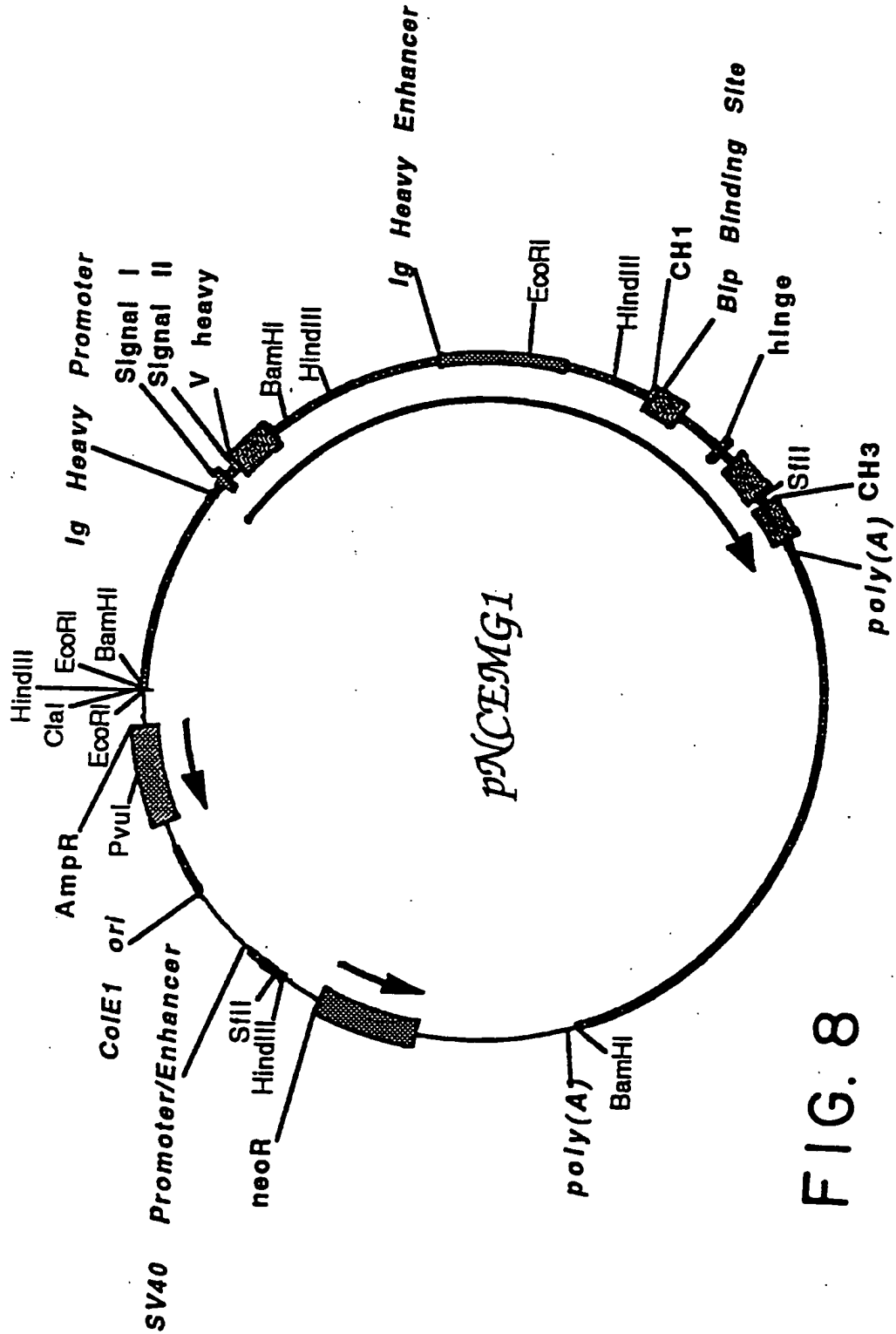
FIG. 8

21 / 22

RBS Primers

RBS 1 B949 (SEQ ID NO: 39)
5` cacacatttctagaAGACACGTACAAACCAatgaaaaagacagcatcgcgattgcagtg 3`

RBS 2 long B849 (SEQ ID NO: 40)
5` cacatttctagaAAGCAAAGTCCCGCGAATGAAAGAGACCTATTTatgaaaagacag
ctatcgcgattgcagtggcactgctggt 3`

RBS 2 short B850 (SEQ ID NO: 41)
5` cacatttctagaAAGCAAAGTCCCGCGAatgaaaagacagcatcgcgattgcagtggcactgctggt 3`

RBS 8 short B844 (SEQ ID NO: 42)
5` cacatttctagaACGTTTAAACAGACACatgaaaagacagcatcgcgattgcagtggcactggctggt 3`

RBS 12 B842 (SEQ ID NO: 43)
5` cacatttctagaGGAACTCAAAGGCACCatgaaaaagacagcatcgcgattgcagtggcactggctggt 3`

Upstream
Primers

Downstream
Primer

940 (SEQ ID NO: 4)
5` aattggtctccggaaccaccgctcaccggctccagat 3`

FIG. 9A

Signal Peptide Primers

Upstream Primers

E10 B946   (SEQ ID NO: 25)
5` taattattctagaaATGACTTCTGGGAGCAAAATAACTGTTCTTATACTTCTTTTG
ACTGGTACACATACAATATATTAGGGggaattccggtcaccagaaacgtcggtgaaagtaaaga 3`

E5 B947   (SEQ ID NO: 24)
5` taattattctagaaATGGGGGATGTACAAGCTGAGACTACTGATAGATTGCCTAAT
GCTTTATGTTTGTTCAACCGATACCggaattccggtcaccagaaacgtcggtgaaagtaaaaga 3`

Downstream Primer

940   (SEQ ID NO: 4)
5` aattggtctcggaaccacgctcaccggctccagat 3`

Note: Letters in bold indicate the sequence taken from the RBS
library clones (between the Xba I site and the ATG start codon
of the signal peptide).

FIG. 9B

A.	CLASSIFICATION OF SUBJECT MATTER

IPC(5)   :C12N 15/63, 15/09; C12P 21/00
US CL   :435/69.1, 69.7, 172.3, 252.3, 320.1; 536/23.4, 24.1
According to International Patent Classification (IPC) or to both national classification and IPC

B.	FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. :   435/69.1, 69.7, 172.3, 252.3, 320.1; 536/23.4, 24.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN/MEDLINE
search terms: random?, promoter#, express?.

C.	DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 4,683,195 (MULLIS ET AL.) 28 July 1987, line 15 of column 29 to line 18 of column 30.	1-72
Y	US, A, 5,096,815 (LADNER ET AL) 17 March 1992, see entire document.	1-72
Y	SCIENCE Vol. 235, issued 16 January 1987, C.A. Kaiser et al., "Many Random Sequences Functionally Replace the Secretion Signal Sequence of Yeast Invertase", pages 312-317, see entire document.	1-27, 39-41, 45-63

[X]	Further documents are listed in the continuation of Box C.	[ ]	See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
05 AUGUST 1994	19 AUG 1994

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	JOHN D. ULM *Jill Warden for*
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	The Journal of Biological Chemistry, Volume 264, issued 05 December 1989, B.D. Lemire et al., "The Mitochondrial Targeting Function of Randomly Generated Peptide Sequences Correlates with Predicted Helical Amphiphilicity", pages 20206-20215, see entire document.	1-27, 39-41, 45-63
Y	NATURE Volume 354, issued 07 November 1991, R.A. Houghten et al., "Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery", pages 84-86, see entire document.	1-27, 39-41, 45-63
Y	Nucleic Acids Research, Volume 16, Number 11, issued 10 June 1988, K.T. Min et al., "Search for the optimal sequence of the ribosomal binding site by random oligonucleotide-directed mutagenesis", pages 5075-5099, see entire document.	28-38, 42-44, 64-72
Y	Nucleic Acids Research, Volume 16, Number 15, issued 11 August 1988, A.R. Oliphant et al., "Defining the consensus sequences of E. coli promoter elements by random selection", pages 7673-7683, see entire document.	28-38, 42-44, 64-72

Billing Information

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Oct. 1 - Oct. 31	Nov. 1, 2004	1 of 9

Previous Balance	$86.77
Payment on Oct. 29	-86.77
Total Current Charges	$0.82
Amount Due by Nov. 27	$80.82

Do not send payment.
The amount below will be billed to your Visa credit card on Nov. 27, 2004

$80.82

Reference P.O. Number:
EB001

Sprint PCS Better Wireless Guarantee.
We are committed to making sure you are completely satisfied with your Sprint PCS wireless experience. We provide a 14-Day Risk-Free Guarantee, Sprint PCS Right Plan Promise, Sprint PCS New For You Upgrade Program and 12-Month Limited Phone Warranty. Visit sprintpcs.com for details.

Get up to 600 Bonus Minutes a year with Sprint 50 at Home.
Receive up to 50 Bonus minutes of domestic Sprint Long Distance on your Home phone a month. That's up to 600 minutes a year! Each additional minute for state-to-state calling is only 7 cents. Sign up for Sprint 50 at Home by calling 1-888-847-4035 or visit www.sprint.com www.sprint.com/bmessage today.

Refer your friends to Sprint, and you'll both save money.
Receive a $20 Sprint PCS Service Credit when a friend you refer to the wireless clarity of Sprint becomes a Sprint PCS customer. Your friends who sign up with Sprint receive a $10 service credit. Restrictions apply. Sign on to sprintpcs.com to learn more about the Sprint PCS Referral Program.

www.sprintpcs.com

You may connect with PCS Customer Solutions by pressing *2 on your PCS phone, *3 to make a one-time payment on your account, or *4 to receive automated account information and change your invoice format. Additional assistance is available by dialing 1-888-211-4727

0005

Sprint PCS

Sprint.

0SPRCHGW
0045090802 88
00002267

A1

CHRISTIAN EHRBAR
39 SALEM ST
WINCHESTER MA 01890-1843

Account Number: 0045090802-8

Do not send payment.
The amount below will be billed to your Visa credit card on Nov. 27, 2004

$80.82

EXPLANATION OF TERMS ON YOUR INVOICE
For additional information, simply visit us at www.sprintpcs.com, sign on to manage your account and click the "My Invoice" button.

Summary of Individual Phone Charges (page 2): A summary of all charges related to each PCS Phone on your account.

- **Monthly Plan/Ad Charges:** The recurring charge for your PCS Service Plans and other services that are invoiced one month in advance.
- **Additional Usage Charges:** This column will show additional voice or roaming minutes of PCS Vision Usage incurred above your minute or usage allowance in your plan, Call Waiting, Three-Way Calling, Voice Command and certain Operator Services.
- **Other Charges** - Charges that are not included in the Monthly Service Charges section will appear here if applicable.
- **Promotions, Credits and Adjustments** - Special offers or pricing promotions appear as credits on your invoice. A displayed charge, when corrected, will appear as an adjustment.
- **Surcharges and Fees** - The surcharges in this section generally recover the costs incurred by Sprint in complying with various federal and state mandates. Charges that appear in this section of your invoice, including charges associated with Federal Wireless Number Pooling and Portability, Federal and State Universal Service Funds (USF) and Federal E911, are neither taxes nor government-imposed assessments. The Federal USF charge is calculated using the FCC-prescribed contribution factor, which may change on a monthly basis, and the federal and state law requires carriers to improve these charges but carriers are permitted to recover their costs of complying with these federal and state mandates.
- **E911** - These include applicable federal, state, city, and county taxes.

Call 1-866-770-8950 for more information, including the current Federal USF invoice surcharge.

*Please note that current availability of E911 service is very limited. E911 service is dependent upon several factors, including the ability of your local public safety agency to receive and process this information and the capabilities of your equipment.

ANYTIME / NIGHT AND WEEKEND MINUTES
Please note that most PCS Service Plans are made up of both Anytime (peak) and Night and Weekend (off-peak) minutes with specific, clearly defined hours for each. The cost of a call is determined by your rate plan and Night and Weekend (cost) typically peak level (when peak) calls made during business hours.

HOW TO REACH PCS CUSTOMER SOLUTIONS
We are making every attempt to simplify your invoice so that it is easy to understand. Still have questions? Here's how to contact us:

Via the Internet:	By Phone:	By Mail:	To Send Payment:
www.sprintpcs.com	By using your PCS Phone; PCS Customer Solutions *2	PCS Customer Solutions	Sprint
	1-888-211-4727 Consumer	P.O. Box 8077	PCS Customer Solutions
	1-888-788-4727 Business	London, KY 40742	P.O. Box 219564
			Kansas City, MO 64121-9564

TTY USERS: ACCESS TO PCS CUSTOMER SOLUTIONS
Please contact a TRS agent (dial 711) and request that he/she dial 1-888-777-4003. PCS Customer Solutions hearing can also be made via the Internet at www.sprintpcs.com.

ADDITIONAL PAYMENT OPTIONS
A number of convenient payment options are available to fit your lifestyle. You may sign up for Auto Pay at www.sprintpcs.com or *3 on your PCS Phone. Depending on credit and payment history, certain payment methods may be restricted.

You may make a one-time payment through www.sprintpcs.com or *3 on your PCS Phone. Depending on credit and payment history, certain payment methods may be restricted.

The Terms and Conditions of PCS Service sometimes change. For the most current version, please visit our Web site at www.sprintpcs.com or dial *1 for PCS Customer Solutions.

This invoice is due and payable upon presentation, and is past due if we do not receive payment by the due date shown on your invoice. Please make your check or money order payable to Sprint in U.S. dollars. Do not send cash. Any payment for Services and equipment not made when due accrues late charges until paid at the rate of 5% per month or at the highest rate allowed by law. Sprint may also begin procedures to terminate service to you. You will be informed of any such termination action as required by law or the Terms and Conditions of Service. If any check sent to Sprint as payment for Services is returned unpaid by your bank, you will be charged $25, or the highest amount allowed by applicable law, for each returned check. Please provide your name and account number on any correspondence to Sprint. We will respond to your letter as soon as possible.

If you have a dispute about any charges on your invoice, you must notify us within 15 days of the date of the invoice or you will be deemed to have accepted the invoice. Initial notice of any dispute may be made by calling PCS Customer Solutions at 1-888-211-4727, for consumers or other written inquiries, please contact PCS Customer Solutions. You may dispute a copy of the Sprint Customer Complaint Resolution procedures by writing one of the Sprint Stores in Puerto Rico or by contacting PCS Customer Solutions. This invoice conditions notice that your service is subject to discontinuation if payment is not received by the due date. California Consumers - To resolve any service or invoicing issues, you should contact PCS Customer Solutions at the toll free number listed above. PCS Service Plans are not regulated by either federal or state regulators. However, should you be unable to resolve any service or banking issues directly with Sprint you may contact the Consumer Affairs branch of the California Public Utilities Commission, 505 Van Ness Avenue, San Francisco, CA 94102 or possibly via TTY.

Sprint is actively working toward solving our nation's solid waste problem. This invoice is printed on recyclable paper.

Account Number _____ **Amount Due** _____

For name changes sign on to www.sprintpcs.com Click on Contact Us> Email us> Enter your Email address> Click on Topic> Choose Other> Enter name change & reason (Marriage, Divorce, Misspelled). For change of address sign on to the above Web site or complete this form. Please PRINT in Blue or Black ink.

MAILING ADDRESS: _____

CITY: _____ STATE: _____ ZIP: _____

HOME PHONE: _____ BUSINESS PHONE: _____

PRRTA

Individual Charges

www.sprintpcs.com

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Oct. 1 - Oct. 31	Nov. 1, 2004	3 of 9

Individual Charges for **CHRISTIAN EHRBAR**
781-696-6990

PCS Free & Clear Plan - Nov. 1 to Nov. 30

Your Plan Includes:
• $49.99 Plan With 500 Included Minutes   • Anywhere On The Nationwide PCS Network
• PCS Add-A-Phone Option   • Caller ID, Call Waiting, Voicemail
• $49.99 Monthly Service Charge   • Three-Way Calling
• 500 Included Minutes Shared for Use

Voice Airtime Summary

Description	Minutes Used
Minutes Used in Plan	241.0
Off-Peak Minutes Used	77.0
Total Voice Minutes	318.0

REMINDER: Minutes are shared on this plan. The "Minutes Used in Plan" shown above in the Voice Airtime Summary only include minutes used on this phone. To determine the shared minute total, add "Minutes Used in Plan" for each phone sharing minutes.

Monthly Service Charges

Description	Charges
PCS Free & Clear Plan	49.99
Long Distance	0.00
PCS Add-A-Phone	10.00
Added Off-Peak Minutes: Evening/Weekend	10.00
	869.99

Additional Usage Charges

Description	Charges
Directory Assistance Charges	1.25
	81.25

Promotions, Credits and Adjustments

Description	Charges
PCS Advantage Agreement Credit	-10.00
	-810.00

Account Summary

www.sprintpcs.com

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Oct. 1 - Oct. 31	Nov. 1, 2004	2 of 9

Summary of
Individual Charges (Individual Usage Summaries begin on page 3)

Individual Information	Monthly Service Charges	Additional Usage Charges & Purchases	Other Charges	Promotions/ Credits/ Adjustments	Taxes */ Surcharges & Fees	Total
CHRISTIAN EHRBAR 781-696-6990	69.99	1.25	0.00	-10.00	7.39	68.63
CHRISTIAN EHRBAR 781-696-6991	10.00	0.00	0.00	0.00	2.19	12.19
Total Current Charges	**79.99**	**1.25**	**0.00**	**-10.00**	**9.58**	**880.82**

* See below for a breakdown of Taxes, and Surcharges & Fees if applicable.

Additional Billing Information
Detail of Taxes, and Surcharges & Fees

Description	Charges
Taxes	
Federal Tax	2.25
Massachusetts State Sales Tax - Services & Usage	3.71
Massachusetts State Wireless 911	0.60
Surcharges & Fees	
Federal Universal Service Fund	1.42
Federal E911	0.80
Federal Wireless Number Pooling And Portability	0.80
	99.58

Individual Charges for **CHRISTIAN EHRBAR** (continued)
781-696-6990

Taxes, and Surcharges & Fees

Description	Charges
Taxes, and Surcharges & Fees	$7.39

Total Individual Charges for CHRISTIAN EHRBAR	$68.63

Need more information?

Visit www.sprintpcs.com for a complete listing of account activity and call detail.

Call Detail

Voice Call Detail

	Date	Time	Phone Number	Call Destination	Rate/ Type	Minutes Used	Airtime Charges	LD/ Additional Charges	Total Charges
1	10/1	5:15 P	617-248-7200	Boston, MA		3.0	included	0.00	0.00
2	10/3	10:10 A	207-829-2389	Cumberland, ME		1.0	included	0.00	0.00
3	10/4	7:45 A	617-248-7200	Boston, MA		1.0	included	0.00	0.00
4	10/4	8:05 P	781-369-1345	Winchester, MA		2.0	included	0.00	0.00
5	10/4	9:50 P	617-248-7200	Boston, MA		3.0	included	0.00	0.00
6	10/5	10:27 A	617-248-7200	Boston, MA		1.0	included	0.00	0.00
7	10/5	10:29 A	617-310-8425	Boston, MA		2.0	included	0.00	0.00
8	10/5	11:03 A	617-248-7200	Boston, MA		2.0	included	0.00	0.00
9	10/5	6:08 P	617-310-8425	Boston, MA		3.0	included	0.00	0.00
10	10/6	6:08 P	617-826-8523	Boston, MA		18.0	included	0.00	0.00
11	10/6	6:24 P	617-248-7200	Boston, MA		1.0	included	0.00	0.00
12	10/7	7:44 A	617-248-7200	Boston, MA		2.0	included	0.00	0.00
13	10/7	7:46 A	617-248-7200	Boston, MA		4.0	included	0.00	0.00
14	10/7	10:13 A	617-248-7200	Boston, MA		4.0	included	0.00	0.00
15	10/7	10:17 A	617-723-9000	Boston, MA		11.0	included	0.00	0.00
16	10/7	10:28 A	617-248-7200	Boston, MA		3.0	included	0.00	0.00
17	10/7	11:27 A	617-248-7200	Boston, MA		1.0	included	0.00	0.00
18	10/7	12:24 P	617-672-7785	Boston, MA		8.0	included	0.00	0.00
19	10/7	12:33 P	617-672-7785	Boston, MA		2.0	included	0.00	0.00
20	10/7	12:35 P	781-369-1345	Winchester, MA		2.0	included	0.00	0.00
21	10/7	12:36 P	781-369-1345	Winchester, MA		2.0	included	0.00	0.00
22	10/7	1:03 P	617-248-7200	Boston, MA		2.0	included	0.00	0.00
23	10/7	1:22 P	617-723-9000	Boston, MA		5.0	included	0.00	0.00
24	10/7	1:27 P	617-248-7200	Boston, MA		1.0	included	0.00	0.00
25	10/7	1:27 P	617-248-7000	Boston, MA		3.0	included	0.00	0.00
26	10/7	1:30 P	781-369-1345	Winchester, MA		1.0	included	0.00	0.00

Individual Charges for **CHRISTIAN EHRBAR** (continued)
781-696-6990

Voice Call Detail

	Date	Time	Phone Number	Call Destination	Rate/ Type	Minutes Used	Airtime Charges	LD/ Additional Charges	Total Charges
27	10/7	3:42 P	781-696-6990	Woburn, MA		3.0	included	0.00	0.00
28	10/7	4:29 P	781-643-0018	Arlington, MA		1.0	included	0.00	0.00
29	10/7	4:34 P	Incoming			18.0	included	0.00	0.00
30	10/7	4:57 P	617-248-7200	Boston, MA		4.0	included	0.00	0.00
31	10/7	8:20 P	781-316-2082	Arlington, MA		1.0	included	0.00	0.00
32	10/7	8:28 P	781-369-1345	Winchester, MA		1.0	included	0.00	0.00
33	10/7	8:30 P	781-316-2082	Arlington, MA		1.0	included	0.00	0.00
34	10/7	8:50 P	781-316-2082	Arlington, MA		3.0	included	0.00	0.00
35	10/7	8:54 P	781-248-1847	Lynn, MA		2.0	included	0.00	0.00
36	10/7	8:57 P	609-818-0739	Pennington, NJ		1.0	included	0.00	0.00
37	10/7	9:45 P	Incoming			20.0	included	0.00	0.00
38	10/7	10:20 P	303-449-1819	Boulder, CO		2.0	included	0.00	0.00
39	10/8	12:17 P	Incoming			2.0	included	0.00	0.00
40	10/8	6:18 P	917-692-4740	New York, NY		11.0	included	1.25	1.25
41	10/8	6:04 P	781-555-1212	Dir. Asst		1.0	included	0.00	0.00
42	10/8	6:05 P	781-861-5098	Lexington, MA		18.0	included	0.00	0.00
43	10/8	6:33 P	781-369-1345	Winchester, MA		6.0	included	0.00	0.00
44	10/8	6:40 P	617-672-7785	Boston, MA		11.0	included	0.00	0.00
45	10/8	12:59 P	Incoming			2.0	included	0.00	0.00
46	10/8	1:29 P	917-692-4740	New York, NY		1.0	included	0.00	0.00
47	10/13	8:19 A	617-248-7200	Boston, MA		1.0	included	0.00	0.00
48	10/13	8:37 P	781-369-1345	Winchester, MA		2.0	included	0.00	0.00
49	10/13	8:40 P	781-369-1345	Winchester, MA		5.0	included	0.00	0.00
50	10/14	7:17 A	617-248-7200	Boston, MA		1.0	included	0.00	0.00
51	10/14	9:00 A	617-248-7200	Boston, MA		2.0	included	0.00	0.00
52	10/14	10:10 P	781-696-6990	Woburn, MA		1.0	included	0.00	0.00
53	10/15	8:44 A	617-248-7200	Woburn, MA		1.0	included	0.00	0.00
54	10/15	8:44 A	617-248-7200	Boston, MA		2.0	included	0.00	0.00
55	10/15	8:48 A	617-672-7785	Boston, MA		3.0	included	0.00	0.00
56	10/15	5:29 P	781-696-6990	Woburn, MA		2.0	included	0.00	0.00
57	10/16	11:28 A	781-369-1345	Winchester, MA		1.0	included	0.00	0.00
58	10/16	1:23 P	781-369-1345	Winchester, MA		2.0	included	0.00	0.00
59	10/17	11:23 A	617-248-7200	Boston, MA		1.0	included	0.00	0.00
60	10/17	12:51 P	617-672-7785	Boston, MA		1.0	included	0.00	0.00
61	10/18	3:38 P	781-369-1345	Winchester, MA		5.0	included	0.00	0.00
62	10/18	6:05 P	617-248-7200	Boston, MA		1.0	included	0.00	0.00
63	10/18	6:07 P	617-248-7200	Boston, MA		1.0	included	0.00	0.00
64	10/19	8:41 A	617-248-7200	Boston, MA		1.0	included	0.00	0.00
65	10/19	11:13 A	Incoming			1.0	included	0.00	0.00
66	10/19	7:12 P	617-248-7200	Boston, MA		2.0	included	0.00	0.00

Individual Charges

www.sprintpcs.com

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Oct. 1 - Oct. 31	Nov. 1, 2004	6 of 9

Individual Charges for CHRISTIAN EHRBAR (continued)
781-696-6990

Voice Call Detail

	Date	Time	Phone Number	Call Destination	Rate/ Type	Minutes Used	Airtime Charges	LD/ Additional Charges	Total Charges
67	10/19	7:14 P B	617-248-7200	Boston, MA		1.0	Included	0.00	0.00
68	10/20	12:01 P B	617-872-7785	Boston, MA		8.0	Included	0.00	0.00
69	10/20	6:15 P B	781-369-1345	Winchester, MA		1.0	Included	0.00	0.00
70	10/21	8:47 A B	617-248-7200	Boston, MA		8.0	Included	0.00	0.00
71	10/21	8:28 P B	617-248-7200	Boston, MA		1.0	Included	0.00	0.00
72	10/21	8:29 P B	617-248-7000	Boston, MA		2.0	Included	0.00	0.00
73	10/21	8:32 P B	781-369-1345	Winchester, MA		2.0	Included	0.00	0.00
74	10/22	8:17 A B	617-248-7200	Boston, MA		1.0	Included	0.00	0.00
75	10/24	3:57 P B	415-393-0024	San Fran, CA		11.0	Included	0.00	0.00
76	10/25	8:26 P B	617-248-7000	Boston, MA		7.0	Included	0.00	0.00
77	10/25	8:38 P B	617-248-7000	Boston, MA		2.0	Included	0.00	0.00
78	10/26	8:53 A B	617-248-7200	Boston, MA		1.0	Included	0.00	0.00
79	10/26	8:42 P B	804-640-9447	Richmond, VA		11.0	Included	0.00	0.00
80	10/27	8:32 A B	617-248-7200	Boston, MA		0.0	Included	0.00	0.00
81	10/27	8:56 A B	617-248-7000	Boston, MA		4.0	Included	0.00	0.00
82	10/27	8:47 P B	617-535-3200	Boston, MA		1.0	Included	0.00	0.00
83	10/27	8:53 P B	617-535-3200	Boston, MA		1.0	Included	0.00	0.00
84	10/27	8:57 P B	781-369-1345	Winchester, MA		1.0	Included	0.00	0.00
85	10/28	8:11 A B	617-248-7200	Boston, MA		2.0	Included	0.00	0.00
86	10/29	7:59 P B	617-723-9000	Boston, MA		12.0	Included	0.00	0.00
87	10/29	8:02 P B	617-723-9000	Boston, MA		3.0	Included	0.00	0.00
88	10/29	8:05 P B	617-248-7200	Boston, MA		2.0	Included	0.00	0.00
89	10/29	8:07 P B	617-248-7200	Boston, MA		1.0	Included	0.00	0.00
90	10/30	3:10 P B	781-696-6591	Woburn, MA		2.0	Included	0.00	0.00
91	10/30	3:54 P B	Incoming			2.0	Included	0.00	0.00
92	10/30	5:40 P B	781-696-6591	Woburn, MA		1.0	Included	0.00	0.00
93	10/30	5:41 P B	781-696-6591	Woburn, MA		1.0	Included	0.00	0.00
94	10/30	5:46 P B	Incoming			1.0	Included	0.00	0.00
95	10/30	5:48 P B	781-696-6591	Woburn, MA		1.0	Included	0.00	0.00
96	10/31	11:08 A B	781-646-4315	Arlington, MA		1.0	Included	0.00	0.00
	Total					86:00	81:25	81:25	81:25

Billing Information

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Sep. 1 - Sep. 30	Oct. 2, 2004	1 of 9

Previous Balance	$239.44
Payment on Sep. 13	-159.04
Payment on Sep. 28	-79.05
Total Current Charges	85.42
Amount Due by Oct. 29	**$86.77**

Do not send payment. The amount below will be billed to your Visa credit card on Oct. 29, 2004

$86.77

Reference P.O. Number: EB001

Sprint takes customer privacy seriously.
Change your Sprint PCS Account Password periodically to keep your customer and call detail information secure. Visit sprintpcs.com today to change your password to one with a minimum of six alphanumeric characters.

Sprint is rewriting the wireless rules.
Making things fair. Making things better. From the new Sprint PCS Fair & Flexible Plan to the Sprint PCS New For You Upgrade Program, we put you first. For more information on our latest service plans and other products and services, visit www.sprintpcs.com.

Access a variety of convenient services when you dial 411 on your Sprint PCS Phone.
Let Sprint PCS Directory Assistance simplify your life. Get phone numbers, theatre locations, driving directions, local event information and more. For only $1.25 per call, plus airtime, get up to three phone numbers or requests answered and get connected to a number at no additional charge.

www.sprintpcs.com

You may connect with PCS Customer Solutions by pressing *2 on your PCS phone. *3 to make a one-time payment on your account, or *4 to receive automated account information and change your invoice format.
Additional assistance is available by dialing   1-888-211-4727 .

Sprint.

0005

Sprint PCS °

SPRINCON
0045090802 00
00223559   A1

CHRISTIAN EHRBAR
33 SALEM ST
WINCHESTER MA 01890-1143

Account Number: 0045090802-8

Do not send payment. The amount below will be billed to your Visa credit card on Oct. 29, 2004

$86.77

Individual Charges

www.sprintpcs.com

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090602-8	Sep. 1 - Sep. 30	Oct. 2, 2004	3 of 9

Individual Charges for **CHRISTIAN EHRBAR**
781-696-6990

PCS Free & Clear Plan - Oct. 1 to Oct. 31
Your Plan Includes:
- $49.99 Plan With 500 Included Minutes
- PCS Add-A-Phone Option
- $49.99 Monthly Service Charge
- 500 Included Minutes Shared for Use

- Anywhere On The Nationwide PCS Network
- Caller ID, Call Waiting, Voicemail
- Three-Way Calling

Voice Airtime Summary

Description	Minutes Used
Minutes Used in Plan	153.0
Off-Peak Minutes Used	71.0
Total Voice Minutes	**224.0**

REMINDER: Minutes are shared on this plan. The "Minutes Used in Plan" shown above in the Voice Airtime Summary only include minutes used on this phone. To determine the shared minute total, add "Minutes Used in Plan" for each phone sharing minutes.

Monthly Service Charges

Description	Charges
PCS Free & Clear Plan	49.99
PCS Add-A-Phone	10.00
Long Distance	0.00
Added Off-Peak Minutes: Evening/Weekend	10.00
	$69.99

Additional Usage Charges

Description	Charges
Voice Long Distance Charges	2.94
Directory Assistance Charges	3.75
	$6.69

Promotions, Credits and Adjustments

Description	Charges
PCS Advantage Agreement Credit	-10.00
	-$10.00

Account Summary

www.sprintpcs.com

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090602-8	Sep. 1 - Sep. 30	Oct. 2, 2004	2 of 9

Summary of Individual Charges (Individual Usage Summaries begin on page 3)

Individual Information	Monthly Service Charges	Additional Usage Charges & Purchases	Other Charges	Promotions/ Credits/ Adjustments	Taxes */ Surcharges & Fees	Total
CHRISTIAN EHRBAR 781-696-6990	69.99	6.69	0.00	-10.00	7.90	74.58
CHRISTIAN EHRBAR 781-696-6991	10.00	0.00	0.00	0.00	2.19	12.19

Summary of Account Charges

Description	Monthly Service Charges	Additional Usage Charges & Purchases	Other Charges	Promotions/ Credits/ Adjustments	Taxes */ Surcharges & Fees	Total
Balance Adjustment On Sep. 25, 2004				-1.25		-1.25
Total Tax Adjustments				-0.10		-0.10
Total Current Charges	**79.99**	**6.69**	**0.00**	**-11.35**	**10.09**	**$85.42**

* See below for a breakdown of Taxes, and Surcharges & Fees if applicable.

Additional Billing Information
Detail of Taxes, and Surcharges & Fees

Description	Charges
Taxes	
Federal Tax	2.42
Massachusetts State Sales Tax - Services & Usage	3.99
Massachusetts State Wireless 911	0.60
Surcharges & Fees	
Federal Universal Service Fund	1.48
Federal E911	0.80
Federal Wireless Number Pooling And Portability	0.80
	$10.09

Individual Charges

www.sprintpcs.com

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Sep. 1 - Sep. 30	Oct. 2, 2004	5 of 9

Individual Charges for CHRISTIAN EHRBAR
781-696-6990

Voice Call Detail

Date	Time	Phone Number	Call Destination	Rate/ Type	Minutes Used	Airtime Charges	LD/ Additional Charges	Total Charges
28	9/10	2:15 P M	617-248-7572	Boston, MA		4.0	included	0.00	0.00
29	9/10	2:19 P M	617-248-7200	Boston, MA		2.0	included	0.00	0.00
30	9/10	2:24 P M	617-555-1212	Dir. Asst		2.0	included	1.25	1.25
31	9/10	2:26 P M	813-830-9244	Ottanhull, OH	LD	8.0	included	2.52	2.52
32	9/10	2:33 P M	617-248-7200	Boston, MA		2.0	included	0.00	0.00
33	9/10	2:35 P M	617-248-7200	Boston, MA		2.0	included	0.00	0.00
34	9/10	4:48 P M	617-248-7200	Boston, MA		1.0	included	0.00	0.00
35	9/10	5:04 P M	704-427-5109	Charlotte, NC		1.0	included	0.00	0.00
36	9/10	5:08 P M	740-427-5109	Gambier, OH		1.0	included	0.00	0.00
37	9/10	5:32 P M	Incoming			5.0	included	0.00	0.00
38	9/12	5:30 A M	781-696-6990	Woburn, MA		3.0	included	0.00	0.00
39	9/12	5:19 P M	Incoming			3.0	included	0.00	0.00
40	9/12	5:31 P M	781-393-1345	Winchester, MA		2.0	included	0.00	0.00
41	9/12	5:33 P M	Incoming			2.0	included	0.00	0.00
42	9/12	5:49 P M	781-393-1345	Winchester, MA		6.0	included	0.00	0.00
43	9/13	8:04 A M	617-248-7200	Boston, MA		1.0	included	0.00	0.00
44	9/13	6:08 P M	617-310-9625	Winchester, MA		2.0	included	0.00	0.00
45	9/13	8:45 P M	781-393-1345	Winchester, MA		5.0	included	0.00	0.00
46	9/14	8:34 P M	212-555-1458	New York, NY		2.0	included	0.00	0.00
47	9/15	8:13 A M	617-248-7200	Boston, MA		1.0	included	0.00	0.00
48	9/15	9:10 A M	781-393-1345	Winchester, MA		1.0	included	0.00	0.00
49	9/15	9:18 A M	781-696-6991	Woburn, MA		5.0	included	0.00	0.00
50	9/15	9:22 A M	781-696-6991	Woburn, MA		2.0	included	0.00	0.00
51	9/15	8:37 A M	617-248-7200	Boston, MA		1.0	included	0.00	0.00
52	9/16	6:02 P M	781-393-1345	Winchester, MA		1.0	included	0.00	0.00
53	9/19	12:44 P M	617-672-7785	Boston, MA		2.0	included	0.00	0.00
54	9/19	2:48 P M	Incoming			6.0	included	0.00	0.00
55	9/19	4:42 P M	Incoming			2.0	included	0.00	0.00
56	9/19	4:45 P M	Incoming			1.0	included	0.00	0.00
57	9/19	5:45 P M	781-755-0307	Winchester, MA		3.0	included	0.00	0.00
58	9/21	7:41 A M	617-535-3200	Boston, MA		2.0	included	0.00	0.00
59	9/21	7:43 A M	617-248-7200	Boston, MA		2.0	included	0.00	0.00
60	9/21	7:45 A M	617-248-7200	Boston, MA		2.0	included	0.00	0.00
61	9/21	10:51 A M	617-248-7200	Boston, MA		4.0	included	0.00	0.00
62	9/21	10:57 A M	617-248-7572	Boston, MA		3.0	included	0.00	0.00
63	9/21	2:48 P M	781-696-6990	Woburn, MA		1.0	included	0.00	0.00
64	9/21	6:04 P M	617-248-7200	Boston, MA		6.0	included	0.00	0.00
65	9/21	6:16 P M	781-393-1345	Winchester, MA		2.0	included	0.00	0.00
66	9/23	6:31 P M	Incoming			11.0	included	0.00	0.00
67	9/23	7:10 P M	781-555-1212	Dir. Asst		2.0	included	1.25	1.25
68	9/23	7:12 P M	781-555-1212	Dir. Asst		2.0	included	1.25	1.25

Individual Charges

www.sprintpcs.com

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Sep. 1 - Sep. 30	Oct. 2, 2004	4 of 9

Individual Charges for CHRISTIAN EHRBAR
781-696-6990

Taxes, and Surcharges & Fees

Description	Charges
Taxes, and Surcharges & Fees	$7.90

Total Individual Charges for CHRISTIAN EHRBAR	$74.58

Need more information?

Visit www.sprintpcs.com for a complete listing of account activity and call detail.

Call Detail

Voice Call Detail

Date	Time	Phone Number	Call Destination	Rate/ Type	Minutes Used	Airtime Charges	LD/ Additional Charges	Total Charges
1	9/1	11:09 A M	617-248-7200	Boston, MA		1.0	included	0.00	0.00
2	9/1	11:10 A M	917-405-8739	New York, NY		3.0	included	0.00	0.00
3	9/1	4:10 P M	617-248-7200	Boston, MA		2.0	included	0.00	0.00
4	9/2	8:47 A M	617-248-7200	Boston, MA		1.0	included	0.00	0.00
5	9/2	11:42 A M	617-248-7200	Boston, MA		1.0	included	0.00	0.00
6	9/2	2:11 P M	617-248-7200	Boston, MA		4.0	included	0.00	0.00
7	9/2	2:51 P M	917-405-8739	New York, NY		2.0	included	0.00	0.00
8	9/2	9:08 P M	617-248-7200	Boston, MA		3.0	included	0.00	0.00
9	9/2	9:19 P M	781-239-7641	Wellesley, MA		1.0	included	0.00	0.00
10	9/2	9:21 P M	781-239-7641	Wellesley, MA		2.0	included	0.00	0.00
11	9/3	3:27 P M	617-248-7200	Boston, MA		2.0	included	0.00	0.00
12	9/4	4:34 P M	781-393-1345	Winchester, MA		2.0	included	0.00	0.00
13	9/4	4:35 P M	781-696-6990	Woburn, MA		3.0	included	0.00	0.00
14	9/4	5:01 P M	781-696-6990	Woburn, MA		2.0	included	0.00	0.00
15	9/4	5:11 P M	781-393-1345	Winchester, MA		2.0	included	0.00	0.00
16	9/7	6:38 P M	781-696-6990	Woburn, MA		2.0	included	0.00	0.00
17	9/7	6:43 P M	781-393-1345	Winchester, MA		1.0	included	0.00	0.00
18	9/9	6:45 A M	781-696-6990	Woburn, MA		1.0	included	0.00	0.00
19	9/9	7:28 P M	781-393-1345	Winchester, MA		1.0	included	0.00	0.00
20	9/10	12:51 P M	617-248-7200	Boston, MA		2.0	included	0.00	0.00
21	9/10	1:41 P M	617-248-7000	Boston, MA		1.0	included	0.00	0.00
22	9/10	1:44 P M	617-248-7200	Boston, MA		3.0	included	0.00	0.00
23	9/10	1:45 P M	781-696-6991	Woburn, MA		2.0	included	0.00	0.00
24	9/10	1:50 P M	310-867-3388	Beverlyhls, CA		2.0	included	0.00	0.00
25	9/10	2:03 P M	Incoming			9.0	included	0.00	0.00
26	9/10	2:12 P M	781-393-1345	Winchester, MA		2.0	included	0.00	0.00
27	9/10	2:14 P M	781-393-1345	Winchester, MA		1.0	included	0.00	0.00

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Sep. 1 - Sep. 30	Oct. 2, 2004	7 of 9

Individual Charges for **CHRISTIAN EHRBAR**
781-696-6991

PCS Add-A-Phone Option - Oct. 1 to Oct. 31

Your Plan Includes:
- $10.00 Monthly Service Charge
- Caller ID, Call Waiting, Voicemail
- Three-Way Calling

Voice Airtime Summary

Description	Minutes Used
Minutes Used in Plan	128.0
Off-Peak Minutes Used	33.0
Total Voice Minutes	**161.0**

REMINDER: Minutes are shared on this plan. The "Minutes Used in Plan" shown above in the Voice Airtime Summary only include minutes used on this phone. To determine the shared minute total, add "Minutes Used in Plan" for each phone sharing minutes.

Monthly Service Charges

Description	Charges
PCS Add-A-Phone Option	10.00
Long Distance	0.00
Added Off-Peak Minutes: Evening/Weekend	0.00
	$10.00

Taxes, and Surcharges & Fees

Description	Charges
Taxes, and Surcharges & Fees	02.19

Total Individual Charges for CHRISTIAN EHRBAR	012.19

Need more information?
Visit www.sprintpcs.com (for a complete listing of account activity and call detail.

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Sep. 1 - Sep. 30	Oct. 2, 2004	6 of 9

Individual Charges for **CHRISTIAN EHRBAR (continued)**
781-696-6990

Voice Call Detail

	Date	Time	Phone Number	Call Destination	Rate/Type	Minutes Used	Airtime Charges	LD/Additional Charges	Total Charges
69	9/23	7:14 P	781-369-1345	Winchester, MA		2.0	included	0.00	0.00
70	9/23	11:51 P	617-248-7200	Boston, MA		2.0	included	0.00	0.00
71	9/25	4:56 P	781-369-1345	Winchester, MA		1.0	included	0.00	0.00
72	9/25	4:58 P	781-369-1345	Winchester, MA		1.0	included	0.00	0.00
73	9/26	2:57 P	617-248-7200	Boston, MA		1.0	included	0.00	0.00
74	9/27	8:50 A	617-872-7785	Boston, MA		7.0	included	0.00	0.00
75	9/27	4:41 P	781-369-1345	Winchester, MA		1.0	included	0.00	0.00
76	9/27	6:34 P	781-369-1345	Winchester, MA		1.0	included	0.00	0.00
77	9/27	7:27 P	781-369-1345	Winchester, MA		1.0	included	0.00	0.00
78	9/28	8:31 A	781-658-6990	Woburn, MA		2.0	included	0.00	0.00
79	9/28	8:33 A	613-783-9605	Ottawahull, ON	LD	1.0	included	0.42	0.42
80	9/28	8:42 A	Incoming			14.0	included	0.00	0.00
81	9/28	6:03 P	617-248-7788	Boston, MA		3.0	included	0.00	0.00
82	9/28	8:21 P	781-658-6691	Woburn, MA		2.0	included	0.00	0.00
83	9/28	6:46 P	781-658-6990	Woburn, MA		1.0	included	0.00	0.00
84	9/28	8:49 P	Incoming			21.0	included	0.00	0.00
85	9/29	6:49 A	617-248-7200	Boston, MA		1.0	included	0.00	0.00
	Total						**$0.00**	**$6.69**	**$6.69**

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Sep. 1 - Sep. 30	Oct. 2, 2004	9 of 9

Individual Charges for   CHRISTIAN EHRBAR
781-696-6991

Voice Call Detail

	Date	Time	Phone Number	Call Destination	Rate/ Type	Minutes Used	Airtime Charges	LD/ Additional Charges	Total Charges
40	9/25	1:09 P	781-369-1345	Winchester, MA		1.0	Included	0.00	0.00
41	9/25	2:36 P	Incoming			4.0	Included	0.00	0.00
42	9/26	1:35 P	781-696-6991	Woburn, MA		2.0	Included	0.00	0.00
43	9/26	3:06 P	781-696-6991	Woburn, MA		1.0	Included	0.00	0.00
44	9/27	9:30 A	781-729-6637	Winchester, MA		6.0	Included	0.00	0.00
45	9/30	9:04 A	617-448-2838	Boston, MA		3.0			0.00
Total							$0.00	$0.00	$0.00

Customer	Account Number	Invoice Period	Invoice Date	Page
CHRISTIAN EHRBAR	0045090802-8	Sep. 1 - Sep. 30	Oct. 2, 2004	8 of 9

Individual Charges for   CHRISTIAN EHRBAR
781-696-6991

Call Detail

Voice Call Detail

	Date	Time	Phone Number	Call Destination	Rate/ Type	Minutes Used	Airtime Charges	LD/ Additional Charges	Total Charges
1	9/4	11:20 A	781-696-6990	Woburn, MA		1.0	Included	0.00	0.00
2	9/6	7:38 P	617-248-7200	Boston, MA		5.0	Included	0.00	0.00
3	9/7	12:04 P	617-310-8106	Boston, MA		2.0	Included	0.00	0.00
4	9/7	1:00 P	781-727-5141	Needham, MA		1.0	Included	0.00	0.00
5	9/8	1:29 P	203-453-5195	Guilford, CT		16.0	Included	0.00	0.00
6	9/8	8:39 P	858-736-4622	La Jolla, CA		1.0	Included	0.00	0.00
7	9/8	8:40 P	858-736-4622	La Jolla, CA		2.0	Included	0.00	0.00
8	9/8	10:18 P	Incoming			12.0	Included	0.00	0.00
9	9/10	10:53 A	617-310-8106	Boston, MA		4.0	Included	0.00	0.00
10	9/10	11:02 A	617-310-8106	Boston, MA		8.0	Included	0.00	0.00
11	9/10	2:01 P	781-696-6991	Woburn, MA		1.0	Included	0.00	0.00
12	9/10	2:02 P	781-696-6990	Woburn, MA		9.0	Included	0.00	0.00
13	9/10	4:00 P	781-369-1345	Winchester, MA		3.0	Included	0.00	0.00
14	9/12	1:21 P	508-653-5728	Natick, MA		1.0	Included	0.00	0.00
15	9/12	1:22 P	781-696-6991	Woburn, MA		2.0	Included	0.00	0.00
16	9/13	4:56 P	781-369-1345	Winchester, MA		1.0	Included	0.00	0.00
17	9/13	3:02 P	781-369-1345	Winchester, MA		1.0	Included	0.00	0.00
18	9/14	8:31 A	617-310-8106	Boston, MA		2.0	Included	0.00	0.00
19	9/14	8:35 A	781-254-4430	Lynn, MA		2.0	Included	0.00	0.00
20	9/14	8:45 A	781-721-2468	Winchester, MA		2.0	Included	0.00	0.00
21	9/14	8:48 A	781-729-7777	Winchester, MA		3.0	Included	0.00	0.00
22	9/14	11:13 A	Incoming			11.0	Included	0.00	0.00
23	9/14	12:16 P	781-696-6991	Woburn, MA		2.0	Included	0.00	0.00
24	9/14	12:35 P	Incoming			13.0	Included	0.00	0.00
25	9/14	12:50 P	Incoming			1.0	Included	0.00	0.00
26	9/14	3:05 P	508-240-5309	Orleans, MA		7.0	Included	0.00	0.00
27	9/14	3:13 P	508-240-5309	Orleans, MA		8.0	Included	0.00	0.00
28	9/14	3:18 P	781-369-1345	Winchester, MA		2.0	Included	0.00	0.00
29	9/14	4:19 P	617-310-8106	Boston, MA		2.0	Included	0.00	0.00
30	9/15	8:16 A	Incoming			16.0	Included	0.00	0.00
31	9/15	8:22 A	Incoming			2.0	Included	0.00	0.00
32	9/16	11:05 A	781-696-6991	Woburn, MA		1.0	Included	0.00	0.00
33	9/16	11:10 A	617-448-2838	Boston, MA		2.0	Included	0.00	0.00
34	9/17	9:16 A	781-696-6991	Woburn, MA		2.0	Included	0.00	0.00
35	9/17	10:49 A	781-696-6991	Woburn, MA		2.0	Included	0.00	0.00
36	9/17	10:51 A	617-310-8106	Boston, MA		4.0	Included	0.00	0.00
37	9/25	12:26 P	781-696-6991	Woburn, MA		2.0	Included	0.00	0.00
38	9/25	12:27 P	781-369-1345	Winchester, MA		2.0	Included	0.00	0.00
39	9/25	12:57 P	617-248-7200	Boston, MA		1.0	Included	0.00	0.00